AC/323(IST-051)TP/207

www.rto.nato.int

RTO TECHNICAL REPORT   **TR-IST-051**

Characterising the Ionosphere

(La caractérisation de l'ionosphère)

Final Report of Task Group IST-051.

Published January 2009

Distribution and Availability on Back Cover

NORTH ATLANTIC TREATY
ORGANISATION

RESEARCH AND TECHNOLOGY
ORGANISATION

AC/323(IST-051)TP/207

www.rto.nato.int

RTO TECHNICAL REPORT

TR-IST-051

Characterising the Ionosphere

(La caractérisation de l'ionosphère)

Final Report of Task Group IST-051.

The Research and Technology Organisation (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote co-operative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective co-ordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also co-ordinates RTO's co-operation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of co-operation.

The total spectrum of R&T activities is covered by the following 7 bodies:

- AVT   Applied Vehicle Technology Panel
- HFM   Human Factors and Medicine Panel
- IST   Information Systems Technology Panel
- NMSG   NATO Modelling and Simulation Group
- SAS   System Analysis and Studies Panel
- SCI   Systems Concepts and Integration Panel
- SET   Sensors and Electronics Technology Panel

These bodies are made up of national representatives as well as generally recognised 'world class' scientists. They also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier co-operation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

Table of Contents

[†] This section was provided as a separate PDF document

Chapter 3 – Mapping

3-1

Chapter 4 – Spontaneous Radio Frequency Emissions from Natural Aurora[†]   4-1

Chapter 5 – Space Weather   5-1

[†] Chapter 4 was provided as a separate PDF document.

Appendix 1 – Proceedings from IST-056 Symposium

List of Figures

Working Group Members

CANADA

Dr. Paul PRIKRYL
Communications Research Centre (CRC)
3701 Carling Avenue
P.O. Box 11490, Station H
Ottawa, Ontario K2H 8S2
Email: paul.prikryl@crc.ca

GERMANY

Dr. Stefan HAWLITSCHKA
FGAN-FKIE
Neuenahrer Strasse 20
53343 Wachtberg
Email: hawi@fgan.de

NORWAY

Dr. Ulf-Peter HOPPE
Norwegian Defence Research Establishment (FFI)
P.O Box 25
NO-2027 Kjeller
Email: uph@ffi.no

UNITED KINGDOM

Prof. Paul S. CANNON
QinetiQ Ltd.
Communications Department
Room PB212
St Andrews Road
Malvern, Worcs WR14 3PS
Email: pscannon@qinetiq.com

Mr. Glyn WYMAN (Chair)
Domain Expert Little Chase
Tidenham Chase, Chepstow NP16 7JN
Email: g.wyman@btinternet.com

UNITED STATES

Dr. Paul BERNHARDT
US Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375
Email: bern@ppd.nrl.navy.mil

Mr. Edward KENNEDY
US Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375
Email: Kennedy@itd.nrl.navy.mil

Characterising the Ionosphere
(RTO-TR-IST-051)

Executive Summary

The NATO RTO Task Group on "Characterising the Ionosphere" was an initiative taken by the Information Systems Technology (IST) Panel to report on the status of research associated with the ionosphere. The aim was to alert the user groups in the wide community, but with an emphasis on the military applications, to the effects of the medium on systems. It also provides guidance on prediction methods for any event which has an influence on the ionosphere. The report is a result of three years work by the Technical Group supported by contributions from academia and other national research institutes. The group actively involved five nations (Canada, Germany, Norway, UK and US) but with grateful contributions from others as appropriate. We were very fortunate to have benefited from the experience of experts in the field; the Group hosted a Specialists' Meeting in Fairbanks Alaska supported by the URSI community; the proceedings are attached as an annex to this report.

The properties of the ionosphere inherently influence the performance of various systems which rely on rf waves traversing the region; for example satellite navigation, HF communications and HF Radar. The resurgence of HF for communications requires improvements to the propagation prediction methods in order to maintain an efficient communications media. Whilst the first order diurnal effects and the influence of space weather are already known, this report describes the physics which could influence the behaviour from events not previously regarded as significant. Several decision aids are described to assist the operators in coping with the inherent instabilities, but identifies further work which would enhance the knowledge. The ionisation is caused by a variety of mechanisms but predominantly by extreme ultra violet (EUV), X-rays, Lyman α and solar cosmic rays. Different ionisation processes occur at specific altitudes giving rise to a layered structure designated as D, E and F; a more comprehensive discussion is available in Chapter 1 "Morphology". The auroral region is a specific case for discussion because it is subject to mechanisms which are generally not well understood. The military community should be aware of the properties and adapt to the effects on their systems when exploiting the media or when deployed in northern latitudes. We have incorporated a section relating to mapping total electron content in near real time, which reduces the latency in establishing the ionospheric propagation delay, a necessary parameter for good navigation using satellite signals e.g. GPS. Galileo Space weather is addressed where interest has been shown from a variety of quarters including the insurance sector.

As chair of the Research Technology Group I am indebted to my colleagues who sat on the technical committee and the scientists who gave presentations to the group. This thanks is extended to the contributors who wrote and presented papers at the Specialists' Meeting/Symposium some of which supported the event through URSI. I would also like to thank the IST Panel and the executive for their continued support throughout the project.

La caractérisation de l'ionosphère
(RTO-TR-IST-051)

Synthèse

Le Groupe Opérationnel de la RTO de l'OTAN sur « La caractérisation de l'ionosphère » est une initiative prise par la commission de Technologie des Systèmes d'Information (IST) afin de faire un rapport sur l'état de la recherche associée à l'ionosphère. Le but était d'alerter les groupes d'utilisateurs dans toute la communauté, mais avec une prépondérance pour les applications militaires, des effets des moyens de communication sur les systèmes. Cela a également donné des conseils sur les méthodes de prévision pour chaque évènement ayant une influence sur l'ionosphère. Le rapport est le résultat de trois ans de travail exécuté par le Groupe Technique soutenu grâce à la contribution de l'académie et d'autres instituts nationaux de recherche. Le Groupe implique activement cinq nations (Canada, Allemagne, Norvège, Royaume Uni et États-Unis) mais avec des contributions très satisfaisantes provenant d'autres pays quand cela était opportun. Nous avons eu la chance de profiter de l'expérience d'experts dans ce domaine ; le Groupe a organisé une Réunion de Spécialistes à Fairbanks en Alaska soutenue par la communauté URSI ; les mesures sont jointes en annexe à ce rapport.

Les propriétés de l'ionosphère influencent fondamentalement la performance de divers systèmes qui reposent sur les ondes radio traversant la zone, comme les satellites de navigation, les communications HF et les radars HF. La résurgence de la HF pour les communications demande d'améliorer les méthodes de prévision de propagation de manière à maintenir des moyens de communication efficaces. Alors que les effets diurnes de premier ordre et l'influence du temps dans l'espace sont déjà connus, ce rapport décrit la physique qui pourrait influencer le fonctionnement à partir d'évènements considérés insignifiants auparavant. Plusieurs aides à la décision sont décrites afin d'aider les opérateurs à prendre en compte les instabilités inhérentes, mais elle identifie un travail supplémentaire qui pourrait augmenter nos connaissances. La ionisation est produite par des mécanismes variés mais principalement par les ultra violets extrêmes (EUV), les rayons X, les α Lyman et les rayons cosmiques solaires. Différents processus de ionisation apparaissent à des altitudes spécifiques donnant naissance à une structure en couches appelées D, E, et F ; des éléments de discussion sont disponibles au chapitre 1 « Morphologie ». La région aurorale est un cas spécifique à débattre car elle est sujette à des mécanismes qui ne sont généralement pas bien compris. La communauté militaire devrait être attentive aux propriétés de l'ionosphère et adapter ces effets sur ses systèmes quand elle utilise ses moyens de communication ou lors d'un déploiement dans les latitudes nord. Nous avons une section qui dresse la carte de la teneur totale en électrons pratiquement en temps réel, ce qui réduit le temps de réaction dans le délai d'établissement de la propagation ionosphérique, c'est un nécessaire paramétrage pour une bonne navigation utilisant les signaux satellites comme le GPS. Nous nous adressons à Galileo Space temps quand un intérêt se présente dans plusieurs directions incluant le secteur assurance.

En tant que responsable du Groupe de Recherches Technologiques, je suis redevable à mes collègues qui siègent au comité technique et aux scientifiques qui font les présentations au groupe. Ces remerciements s'adressent aussi aux collaborateurs qui ont écrit et présenté des documents à la Réunion des Spécialistes/ Symposium, quelques-uns d'entre eux en tant que membre d'URSI. Je voudrais aussi remercier la commission IST et la direction pour son soutien durant tout le projet.

Chapter 1 – MORPHOLOGY AND DYNAMICS

1.1   GENERAL INTRODUCTION

1.1.1   The Ionosphere

(Stefan Hawlitschka, FGAN Germany)

The Ionosphere is a constituent of the atmosphere where at altitudes in excess of 50 km the earth's atmosphere becomes increasingly rarefied and can support free electrons for short periods. This region has the properties of a gas and a plasma which exhibit some interesting effects. Neutral atoms and molecules, such as N_2, N, NO, O_2, O, H and He are ionised by impact of photons of the *extreme ultraviolet* (EUV) and *X-ray* radiation, emanating from the sun. Different mechanisms are predominant at specific altitudes; at the outer reaches of the atmosphere the solar radiation amounts to some 1370 Watts per square metre but the density of atoms is low resulting in a small amount of ionisation. At lower altitudes atoms are more numerous increasing the rate of ionisation but at the same time allowing increased rate of recombination. A further mechanism is the absorption of the incident energy which is a function of the gas density and composition; causing attenuation to increase at lower altitudes. The resultant effect is horizontal stratification denoted as the *D-*, *E-*, and *F-layer* of the ionosphere (Figure 1-1). The *C-layer* also exists generated by cosmic rays but plays a minor role for our purposes.
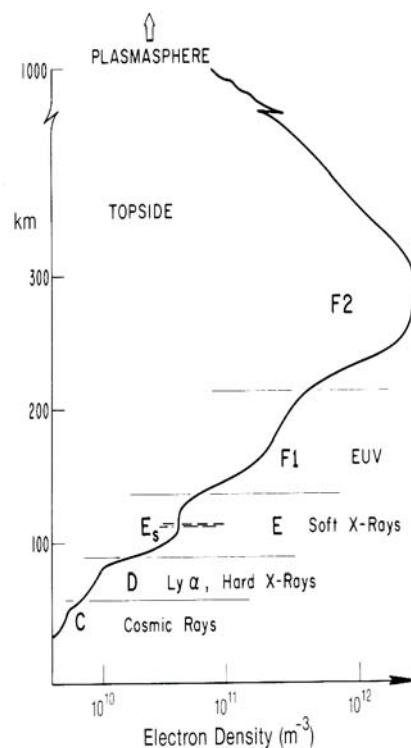
Figure 1-1: Ionosphere Structure on a Summer Day in a Middle Latitude, and the Main Bands of Solar and Cosmic Ionizing Radiation – from Davies [4].

Electromagnetic *high frequency* (HF) waves propagating through the ionosphere may experience reflection, refraction, attenuation and modification of the plane of polarisation. Reflection of HF waves occurs depending on their frequency, the degree of ionisation of the layers of the ionosphere, and on the angle of propagation. The smaller the elevation of the HF wave the higher is the maximum frequency that

can be reflected. As characteristic parameter the *critical frequency* is specified which denotes the highest frequency that is being reflected at vertical incidence. It is subject to high variability and depends on the production rate of free electrons, the loss rate which is mainly due to recombination and the transport of the charged particles.

The ionosphere is not only controlled by the radiation of the sun and the coupling with the neutral atmosphere, but also affected by the solar wind, the earth's magnetic field as well as the interplanetary magnetic field and other phenomena. The resulting high variability of the density of charged particles impedes a detailed forecast of the actual ionosphere. Modelling is therefore mostly constrained to predict in general terms only, average states of the ionosphere. The *International Reference Ionosphere* (IRI) is a climatological model describing the ionospheric parameters for quiet conditions, which is subject to constant review and has recently been modified to cover disturbed conditions. It is recognised that describing the necessary parameters to characterise a storm is not a trivial exercise and knowledge of the heterogeneities of the ionosphere an important and challenging task.

1.1.1.1   The Layers of the Ionosphere

The primary process to contribute to the production of free charge carriers is *photoionization* of gases by the solar extreme ultraviolet (EUV) and X-ray radiation. At polar latitudes, precipitation of energetic particles from the solar wind also plays an essential role in ionization production. Simultaneously a *recombination process* counters the ionisation Therefore at night, the lower layers D and E, which lie in a comparatively dense atmosphere, disappear while the F-layer persists because of its greater height where the atmosphere exhibits a smaller density and thus the collision frequency of the particles is low. At dawn the production of charged particles is restarted and the free electron density N_e increases until the production rate equals the recombination rate.

The lowest noticeable ionospheric region (or layer) is the D region between 50 km and 90 km above ground. The recombination rate is high in this region and as a result the free electron and ion densities are small. Electron density is not high enough for HF waves to be reflected, but electromagnetic waves may be attenuated. Typical day-time densities of free electrons lie around 10^9 to 10^{10} electrons/m^3. The D region appears sun-synchronously and nearly disappears at night. During a radiative eruption from the sun's surface, known as a solar flare, a sudden ionospheric disturbance (SID) can occur. A great enhancement of solar X-rays which penetrate deeply into the atmosphere causes much higher ionization of the D region. It may occur within a few seconds and result in strong absorption of HF waves and interrupt of communication circuits (Mögel-Dellinger effect). The ionization decreases again and returns to its normal density in about an hour.

The E-layer is located between 90 km and 120 km above ground. The electron densities may exceed 10^{11} electrons/m^3 during the day while at night densities are around 3×10^9 electrons/m^3. The daytime electron and ion concentrations are only about 10^{-11} of the neutral gas. Thus the structure of the plasma is strongly influenced by processes in the neutral atmosphere. Slant HF waves up to 10 MHz are reflected by this layer, the daytime critical frequency foE (vertical incidence) is about 4 MHz while during the night foE reaches a minimum of about 0.5 MHz. The E-region exists predominantly in daytime and begins to disappear at dusk due to recombination. The seasonal variations are in phase with solar elevation, but the critical frequency foE varies according to season but is less than 1 MHz. The solar cycle dependent variatіon of foE is about 0.5 MHz.

A sporadic E layer can be observed, which as the name implies is not always present The mechanisms of generation are attributed to thunderstorms and/or wind shears but are not yet fully understood and is not predictable. Electron densities can reach 10^{14} electrons/cubic metre and is significant in that it will reflect waves up to a frequency of approximately 200MHz.

The F-layer is the most dense regular and the most important ionospheric region regarding the influence onto the propagation of electromagnetic waves. Electron densities are of the order of 10^{12} electrons/m³. The daytime electron and ion concentrations are about 10^{-5} of the neutral gas which again strongly influences the structure of the plasma. But due to the higher electron and ion densities the impact of the magnetic field is stronger than in the E-region. At heights of about 300-400 km the gravitational and the magnetic forces are equal. The F-region extends between 120 km and 1000 km above ground with a maximum of the ionization in the range of 200 km to 400 km. It can split up into the F_1 and F_2 layer. The F_1 layer is observed only during daytime. It is more pronounced during summer than winter, at low sunspot numbers, and during ionospheric storms. It's observed best, when f_oF_2 is low. The seasonal variation of the critical frequency f_oF_1 is less than 1 MHz. The solar cycle dependent variation of f_oF_1 can reach 2 MHz.

The most variable of the regular ionospheric layers is the F_2 layer. Whereas variations of f_oE and f_oF_1 are comparatively regular, i.e., diurnal and seasonal, and in phase with the elevation of the sun, the maximum f_oF_2 at noon maybe twice in winter than in summer. The peak electron density has thus to be four times higher in winter than in summer what is called the *winter anomaly* and occurs in daytime only. The diurnal variation of f_oF_2 is much bigger in winter than in summer. During the maximum of the solar cycle the maximum noon f_oF_2 can reach twice the value during the solar minimum. The F_2 layer is influenced by the geomagnetic field. This can be seen in the fact that the worldwide distribution of the critical frequency is aligned with the magnetic equator.

At heights above about 1000 km the atmosphere is fully ionised and therefore the earth's magnetic field controls the structure of the plasma.

The critical frequencies of all layers grow with increasing sunspot numbers. The variability of the F_2 layer with the 11-year sunspot cycle is the highest among the regular layers. In Figure 1-2, the seasonal as well as the solar cycle variations are seen. The seasonal variation of f_oE and f_oF_1 are in phase with solar elevation whereas f_oF_2 is in antiphase (winter anomaly). The F_1 lar disappears in winter. There is a marked increase of the critical frequencies with the 11-year sunspot cycle [4].
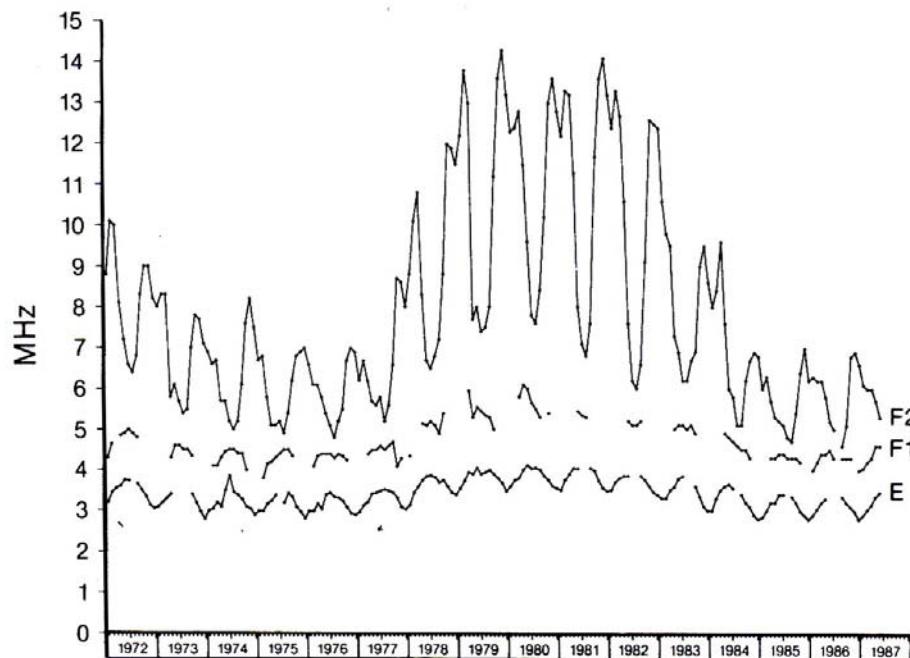
Figure 1-2: Monthly Median Noon Critical Frequencies at Washington DC, from 1972 to 1987 – from Davies [4].
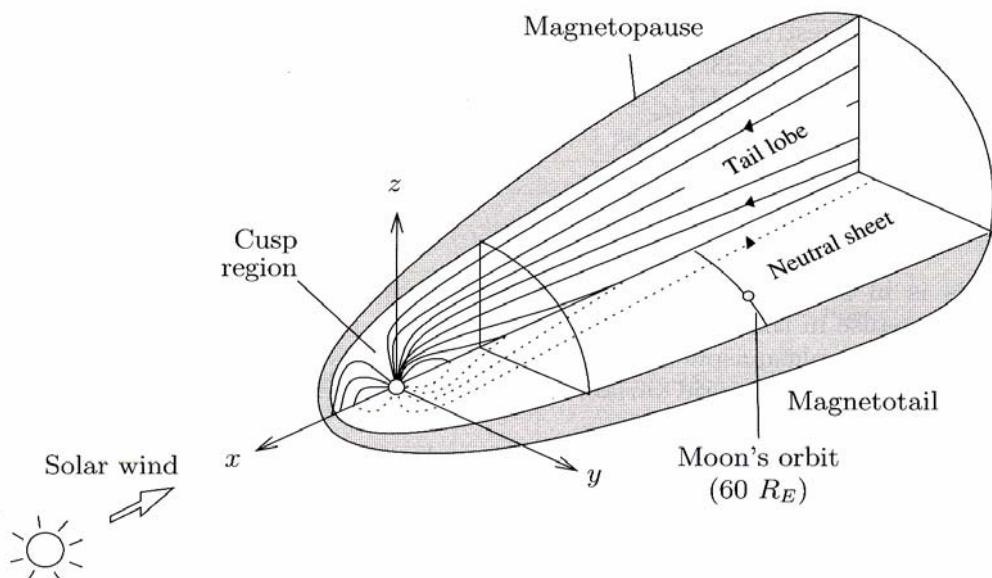
Figure 1-3: General Configuration of the Magnetosphere – from Proelss [12].

The dipolar terrestrial magnetic field is distorted by the interaction with the solar wind and the *frozen-in* interplanetary magnetic field (IMF) that significantly modify the structure of the terrestrial magnetic field. As a result, the earth magnetic field is confined to a finite volume of asymmetric shape. The outer boundary is called magnetopause. On the sunward dayside it assumes an ellipsoidal shape whereas on the opposite side it is characterized by a long cylindrical tail reaching far into space. The cusp region denotes a region of weak magnetic field where the magnetosphere is "open" for the input of the solar particle flux.

1.1.1.2   Disturbances

The ionosphere interacts with solar events, the terrestrial and interplanetary magnetic field as well as with the neutral atmosphere. Solar flares are characterized by short-duration (? 1h) strong enhancement of solar radiation. The radiation flux enhancement in the EUV ranges up to a factor of two and in the X-rays ranges up to a factor of 10^4 and has a significant impact upon the ionosphere. The occurrence frequency of solar flares generally follows the rhythm of the solar cycle, whereby about 5 major events occur every year when the solar activity is low and 70 such events are recorded yearly around the solar maximum.

Ionospheric storms result from very large energy inputs from the sun to the polar atmosphere and the interaction with *geomagnetic storms*. During *coronal* or *solar mass ejections* (CME/SME) a huge amount of solar material is ejected from the sun which propagates through interplanetary space with *frozen-in* magnetic fields forming *magnetic clouds*. When they impinge on the terrestrial magnetic field, initial conditions for the rise of a geomagnetic storm are given, especially when the interplanetary magnetic field is pointing southward (negative B_z). During a solar mass ejection magnetic clouds consisting of plasma are being ejected into space, and strong negative B_z values can occur. The resulting changes of the terrestrial magnetic field are described as *magnetic activity* whose strength is given by the *Dst* (disturbance storm) index or the three hourly local K- or the planetary K_p-indices. The terrestrial magnetic field deviates the charged particle flux around the earth. Over the poles the particles may travel along magnetic field lines and penetrate into the *magnetosphere* including the *magnetotail*. An *ionospheric storm* is triggered. One would expect that the input of particles leads to an increase of the ionization of the ionosphere. However at mid-latitudes both an increase as well as a decrease is observed. The former is referred to as a *positive ionospheric storm* while the latter as a *negative ionospheric storm*. A negative storm occurs when the atmospheric (thermospheric) gas composition changes during the storm. The density of atomic oxygen

decreases at F_2 layer heights and thus the oxygen ion production rate decreases too, whereas the N_2 density increases and, as a result, the recombination rate of the respective ions increases. A positive storm is triggered by traveling *atmospheric* disturbances (TAD) at high latitudes. The ionized component of the traveling atmospheric disturbance, the *traveling ionospheric disturbance* (TID) is deviated by the magnetic field to higher heights and with it the F-region is lifted. The production rate of ions and free electrons decreases more slowly with height than the loss coefficient. Thus the densities of charged particles are augmented. Many questions concerning the generation and the propagation of TIDs are not yet fully understood. But it is generally accepted that during high magnetic activity at high latitudes the atmosphere is being heated during particle precipitation (Joule heating). The friction between the drifting charge carriers and the neutral gas particles heats up the thermospheric gases. TADs with an average propagation velocity of 600 m/s (2160 km/h) are generated during *substorms*, the ionized component can be detected at mid-latitudes as large-scale TIDs which arrive at equatorial latitudes after about 4 h. The rise of a substorm during a storm phase is complicated and is connected to currents in the tail of the magnetosphere. The density disturbance propagation towards the equator is called convective disturbance transport. It is usually stronger on the night-side part of the earth because on the day-side winds are predominantly poleward directed.

As stated above a common feature of *large-scale* TIDs that are generated in the auroral zone is that they are the ionospheric signature of internal gravity waves which have been triggered by sudden expansions of the neutral gas density. Large-scale TIDs are characterised by horizontal wavelengths of up to a few thousands kilometer and periods of about half an hour up to about 4 hours. They are connected with winds propagating equatorwards at velocities of 50-200 m/s (180-720 km/h).

Medium-scale atmospheric gravity waves are oscillations of the neutral atmosphere which can be observed by their traces in the ionospheric plasma. MS Gravity waves cause medium-scale traveling ionospheric disturbances (MSTIDs), which are manifested through changes of the local plasma density, plasma temperature and field-aligned plasma velocity. Changes of the plasma density have been observed up to 40% (Bristow and Greenwald 1995). They estimated gravity wave parameters by estimation of oblique backscatter and its variations due to TIDs using a HF Radar. The wave amplitudes are inferred by comparing the HF radar observations with simulations.

1.1.1.3   Spread-F and Sporadic-E

Mid-latitude spread-F is identified in ionograms by multiple and spread F layer traces. These are caused by strong undulations and irregular plasma structuring in the F region electron density (Haldoupis et al. 2003). In the literature several generation mechanisms are described, ranging from the passage of large scale atmospheric gravity waves, electrodynamic forces and large-scale plasma instabilities. Haldoupis et al. (2003) postulated that spread-F can be generated from eastward electric fields and that there is a link between meso-scale spread-F and unstable sporadic-E layers (E_s).They present a mechanism spread-F generation in which the key role is played by polarization electric fields set up within westward-drifting E_s plasma patches and mapping up to the F region.

1.1.2   References

[1]   Afraimovich, E. L., Kosogorov, E. A., Lesyuta, O. S., Ushakov, I. I., Yakovets, A. F. (2001): *Geomagnetic Control of the spectrum of travelling ionospheric disturbances based on data from a global GPS Network*. Ann. Geophys. 19, 723-731.

[2]   Bristow, W. A. and R. A. Greenwald (1995): *Estimating gravity wave parameters from oblique high-frequency backscatter: Modeling and analysis*. J. Geophys. Res., 100(A3), 3639-3648.

[3]   Budden, K. G. (1985): *The propagation of radio waves*. Cambridge University Press, ISBN 0 521 36952 2.

[4]   Davies, K. (1990): *Ionospheric Radio*. Peter Peregrinus Ltd, ISBN 0 86341 186 X.

[5]   Francis, S. H. (1975): *Global propagation of atmospheric gravity waves : A review*. JASTP 37, 1011-1054.

[6]   Haldoupis, C., M. C. Kelley, G. C. Hussey and S. Shalimov (2003): *Role of unstable sporadic-E layers in the generation of mid-latitude spread F*. J. Geophys. Res., 108(A12), 1446, doi:10.1029/ 2003JA009956.

[7]   Hawlitschka, S. (2006a): *Travelling ionospheric disturbances (TIDs) and tides observed by a super-resolution HF direction finding system*. JASTP 68 (2006) 568-577.

[8]   Hawlitschka, S. (2006b): *Direction finding errors induced by plasmawaves of the ionosphere. Specialists Meeting/Symposium on Characterising the Ionosphere*, 12-16 June 2006, Fairbanks, USA.

[9]   Hocke, K., Schlegel, K. (1996): *A review of atmospheric gravity waves and travelling ionospheric disturbances: 1982-1995*. Ann. Geophys. 14, 917-940.

[10]   Hunsucker, R. D. (1982): *Atmospheric Gravity Waves Generated in the High-Latitude Ionosphere: A Review*. Reviews of Geophysics and Space Physics, Vol. 20, No. 2, 293-315.

[11]   Kalikhman, A. D., (1980): Medium-scale travelling ionospheric disturbances and thermospheric winds in the F-region. JASTP 42, 697-703.

[12]   Proelss, G. W. (2004): *Physics of the Earth's Space Environment*. Springer, ISBN 3-540-21426-7.

[13]   Samson, J. C., R. A. Greenwald, J. M. Greenwald, J. M. Ruohoniemi, A. Frey and K. B. Baker (1990): *Goose Bay Radar Observations of Earth reflected, Atmospheric Gravity Waves in the high-Latitude Ionosphere*. JGR, Vol. 95, NO. A6, 7693-7709.

[14]   Tsybulya, K. and N. Jakowski (2005): *Medium- and small-scale irregularities detected by GPS radio occultation method*. Geophysical Research Letters, 32 (9).

[15]   Waldock, J. A., Jones, T. B. (1987): *Source regions of medium-scale travelling disturbances observed at mid-latitudes*. JASTP 61, 629-639.

[16]   Yakovets, A. F., Kaliev, M. Z., Vodiannikov, V. V., (1999): *An experimental study of wave packets in travelling ionospheric disturbances*. JASTP 61, 629-639.

1.2   MID-LATITUDE

1.2.1   Neutral Atmosphere

(Ulf-Peter Hoppe, Norwegian Defence Research Establishment (FFI))

The ionosphere is embedded in the neutral atmosphere. Even at the maximum of the F-region, with a typical electron number density of 10^{12} m^{-3}, the neutral number density is about 3×10^{14} m^{-3}. At lower altitudes, in the E- and D-regions, the neutral number density overwhelms the electron number density by as much as 10 orders of magnitude. For this reason, neutral atmosphere processes such as the general circulation, tides, planetary and gravity waves as well as the chemical composition down to the number densities of source constituents like O_3 and NO sometimes play important roles for the characteristics of the ionosphere (e.g., Fuller-Rowell et al., 1996).

The neutral atmosphere is described in a statistical sense by model atmospheres such as the COSPAR International Reference Atmosphere (CIRA-86; Fleming et al., 1988; Fleming et al., 1990). For some altitudes, latitudes, and for some seasons, better statistical datasets are available (e.g., Lübken, 1999; Lübken and von Zahn, 1991). Natural fluctuations in all parameters of the atmosphere make individual observations differ from any of these average descriptions.

A number of recent studies address long-term trends in the upper atmosphere and ionosphere. Rishbeth (1997) pointed out how essential it was to maintain a basic world-wide programme of upper atmospheric and geomagnetic monitoring in order to detect any such changes. Lean (1997) came to a similar conclusion for the sun-earth system. A nearly linear decrease in the altitude of the F_2-layer peak over 39 years at Sodankylä, Finland, is indicative of a greenhouse cooling of the upper atmosphere and ionosphere, as is to be expected together with a greenhouse warming in the troposphere (Ulich and Turunen, 1997; Ulich et al., 2003). Cannon et al. (2004) have shown that a 16 km drop in $h(m)F_2$ will introduce ground errors of similar to 100 km. They point out an urgent need to reach a consensus understanding of long-term ionospheric trends. A reassessment of the CCIR and URSI worldwide maps of ionospheric characteristics is likely necessary. Hall and Cannon (2002) suggest that the greenhouse cooling at high altitudes is at least partly causal for the observed negative trend in the critical frequency of the F_2-layer over Tromsø. Bremer et al. (2004) argue that these findings are not global in nature, but limited to individual stations. These authors point out that in the F_1-layer the derived mean global trend in f_oF_1 is in good agreement with model predictions of an increasing greenhouse effect. The trend results in the D-region derived from ionospheric reflection height and absorption measurements in the LF, MF and HF ranges can at least partly be explained by an increasing atmospheric greenhouse effect.

Sprites and elves are recently discovered manifestations of neutral atmosphere-ionosphere interactions (e.g., Pasko et al., 2002; Moudry et al., 2003; Valdivia, 2003; Tong et al., 2004).

Tidal motions and planetary waves in the thermosphere have significant influence on ionospheric properties (Meyer and Forbes, 1997). For instance, the semi-diurnal and quarterdiurnal tides cause ionisation convergence into layers in the equatorial anomaly region of the lower ionosphere (Lee et al., 2003). Planetary waves propagating upward from the stratosphere and mesosphere influence the E- and F-regions (Lawrence and Jarvis, 2003) by transport and mixing.

The neutral atmosphere being 100 times to 10^{10} times more massive than the ionosphere, the former influences the latter much more than vice versa (e.g., Titheridge, 1995). However, the magnetospherically imposed ion convection pattern actually sets up a neutral convection pattern in the thermosphere which in turn modifies the distribution of neutral tracer constituents (Burns et al., 2004). Aruliah et al. (2004) report that the magnitude of the upper thermosphere neutral wind dynamo field is on average 50% of the magnetospheric electric field.

Some of the well-known F_2-layer anomalies might have their origins in the lower atmosphere. It has been suggested that phenomena such as storms, earthquakes and volcanic explosions may produce F_2-layer signatures (Rishbeth, 2006). Kazimirovsky et al. (2003) have reviewed such possible effects. Pulinets et al. (2006) have published a case study of anomalous variations of the total electron content (TEC) registered over the epicenter of an impending earthquake three days prior to the main earthquake event. Calais and Minster (1998) have demonstrated the GPS capabilities to detect the ionosphere perturbations immediately after an earthquake. Underground nuclear explosions emit infrasonic acoustic waves that propagate even to the ionosphere and may perturb the electron number density by as much as 10% (Krasnov and Drobzheva, 2005). Pokhotelov et al. (1995) have reviewed acoustic-ionosphere coupling in general. Aburjania and Machabeli (1998) have examined the coupling of infrasound waves into the ionosphere and mention earthquake precursors in this context.

Ice particles in the mesosphere, such as in noctilucent clouds (NLC) and Polar Mesosphere Summer Echoes (PMSE) modify the D-region at high latitudes in summer quite dramatically (e.g., Gumbel et al., 2003) – see also the section on PMSE below.

1.2.2   References

Aburjania, G.D., Machabeli, G.Z., Generation of electromagnetic perturbations by acoustic waves in the ionosphere, J. Geophys. Res. 103 (A5), 9441-9447, 1998.

Aruliah, A.L., Griffin, E.M., McWhirter, I., Aylward, A.D., Ford, E.A.K., Charalambous, A., Kosch, M.J., Davis, C.J., Howells, V.S.C., First tristatic studies of meso-scale ion-neutral dynamics and energetics in the high-latitude upper atmosphere using collocated FPIs and EISCAT radar, Geophys. Res. Lett. 31 (3), Art. No. L03802, 2004.

Bremer J., Alfonsi, L., Bencze, P., Lastovicka, J., Mikhailov, A.V., Rogers, N., Long-term trends in the ionosphere and upper atmosphere parameters, Ann. Geophys. 47 (2-3), 1009-1029, 2004.

Cannon, P.S., Rogers, N.C., and Hall, C.M., Trends in critical frequencies and layer heights over Tromsø and their consequential impact for radio system modeling, IRI: Quantifying Ionospheric Variability, Adv. Space Res. 34 (9), 2085-2091, 2004.

Burns, A.G., Wang, W., Killeen, T.L., and Solomon, S.C., A "tongue" of neutral composition, J. Atmos. Solar-Terr. Phys. 66 (15-16), 1457-1468, 2004.

Calais, E., and Minster, J.B., GPS, earthquakes, the ionosphere, and the Space Shuttle, Phys. Earth Planet. Interiors 105 (3-4), 167-181, 1998.

CIRA-86, http://badc.nerc.ac.uk/data/cira/.

Fleming, E.L., Chandra, S., Shoeberl, M.R., and Barnett, J.J., Monthly mean global climatology of temperature, wind, geopotential height and pressure for 0-120 km, National Aeronautics and Space Administration, Technical Memorandum 100697, Washington, D.C., 1988.

Fleming, E.L., Chandra, S., Barnett, J.J., and Corney, M., Zonal mean temperature, pressure, zonal wind, and geopotential height as a function of latitude, Adv. Space Res., 10(12), 11-59, 1990.

FullerRowell, T.J., Codrescu, M.V., Rishbeth, H., Moffett, R.J., Quegan, S., On the seasonal response of the thermosphere and ionosphere to geomagnetic storms, J. Geophys. Res. 101 (A2), 2343-2353, 1996.

Gumbel, J., Siskind, D.E., Witt, G., Torkar, K.M., Friedrich M., Influences of ice particles on the ion chemistry of the polar summer mesosphere, J. Geophys. Res. 108 (D8), Art. No. 8436, 2003.

Hall, C.M. and Cannon P.S., Trends in f_oF_2 above Tromso (69 degrees N 19 degrees E), Geophys. Res. Lett. 29 (23), Art. No. 2128, 2002.

Kazimirovsky, E., Herraiz, M., De la Morena, B.A., Effects on the ionosphere due to phenomena occurring below it, Surveys Geophys. 24 (2), 139-184, 2003.

Krasnov, V.M. and Drobzheva, Y.V., The acoustic field in the ionosphere caused by an underground nuclear explosion, J. Atmos. Solar-Terr. Phys. 67 (10), 913-920, 2005.

Lawrence, A.R. and Jarvis, M.J., Simultaneous observations of planetary waves from 30 to 220 km, J. Atmos. Solar-Terr. Phys 65 (6), 765-777, 2003.

Lean, J., The Sun's variable radiation and its relevance for Earth, Ann. Rev. Astron. Astrophys. 35, 33-67, 1997.

Lee, C.C., Liu, J.Y., Pan, C.J., Hsu, H.H., The intermediate layers and associated tidal motions observed by a digisonde in the equatorial anomaly region, Ann. Geophys. 21 (4), 1039-1045, 2003.

Lübken, F.-J., Thermal structure of the Arctic summer mesosphere, J. Geophys. Res., 104, 9135-9149, 1999.

Lübken, F.-J. and U. von Zahn, Thermal structure of the mesopause region at polar latitudes, J. Geophys. Res. 96, 20, 841-20, 857, 1991.

Pasko, V.P., Stanley, M.A., Mathews, J.D., Inan, U.S., Wood, T.G., Electrical discharge from a thundercloud top to the lower ionosphere, Nature 416 (6877), 152-154, 2002.

Meyer, C.K., Forbes, J.M., Natural oscillations of the ionosphere-thermosphere-mesosphere (ITM) system, J. Atmos. Solar-Terr. Phys. 59 (17), 2185-2202, 1997.

Moudry, D., Stenbaek-Nielsen, H., Sentman, D., Wescott, E., Imaging of elves, halos and sprite initiation at 1 ms time resolution, J. Atmos. Solar-Terr. Phys. 65 (5), 509-518, 2003.

Pokhotelov, O.A., Parrot, M., Fedorov, E.N., Pilipenko, V.A., Surkov, V.V., Gladychev, V.A., Response of the Ionosphere to Natural and Man-Made Acoustic Sources, Ann. Geophys. 13 (11), 1197-1210, 1995.

Pulinets, S.A., Ouzounov, D., Ciraolo, L., Singh, R., Cervone, G., Leyva, A., Dunajecka, M., Karelin, A.V., Boyarchuk, K.A., Kotsarenko, A., Thermal, atmospheric and ionospheric anomalies around the time of the Colima M7.8 earthquake of 21 January 2003, Ann. Geophys. 24 (3), 835-849, 2006.

Rishbeth, H., Long-term changes in the ionosphere, Middle Atmosphere: Changes and Electrodynamics, Advances in Space Research 20 (11), 2149-2155, 1997.

Rishbeth, H., F-region links with the lower atmosphere? J. Atmos.Solar-Terr.Phys. 68 (3-5), 469-478, 2006.

Titheridge, J.E., Winds in the Ionosphere - A Review, J. Atmos. Terr. Phys. 57 (14), 1681-1714, 1995.

Tong, L.Z., Nanbu, K., Hiraki, Y., Fukunishi, H., Particle modeling of the electrical discharge in the upper atmosphere above thundercloud, J. Phys. Soc. Japan 73 (9), 2438-2443, 2004.

Ulich, T. and E. Turunen, Evidence for long-term cooling of the upper atmosphere in ionosonde data, Geophys. Res. Lett. 24 (9), 1103-1106, 1997.

Ulich, T., Cliverd, M.A., Rishbeth, H. Determining the long-term change in the ionosphere. EOS Trans. 84(52), 581-585, 2003.

Valdivia, J.A., Lightning induced optical emissions in the ionosphere, Space Sci. Rev. 107 (1-2), 273-291, 2003.

1.3   HIGH-LATITUDE IONOSPHERE

1.3.1   Overview of Important Effects

(Paul Prikryl, Communications Research Centre Canada)

The high-latitude ionosphere differs significantly from its mid-latitude counterpart. The basic processes of ionization production, loss and transport that shape the ionosphere at mid latitudes also operate at high latitudes thus producing main ionospheric layers as described above. However, given the geometry of the geomagnetic field which is to the first approximation dipolar, the high-latitude ionosphere is modified by dynamic magnetospheric processes that are largely controlled by the interaction between the solar wind and the earth's magnetosphere. In fact, many ionospheric phenomena that occur at high latitudes are "footprint signatures" of this interaction and the ionosphere is sometime likened to a nature's TV screen for viewing the images of auroral displays. Through the coupling to the interplanetary magnetic field (IMF) the polar ionosphere gains direct access to solar wind plasma and particles that have been energized by magnetospheric processes and are sources of ionization in the auroral region in particular. By the same token, the ionospheric plasma can escape to space (polar wind and auroral bulk upflows of ions with energy of a few eV) or be trapped in the magnetosphere. A fraction of up-flowing ions can be accelerated to high energies through various processes that include ion beams, ion conics, upwelling ions and transversely accelerated ions.

1.3.1.1   Ionization Sources

At high latitudes, the main source of ionization, which is solar EUV and X rays, is supplemented by energetic particle precipitation. The electrons that are accelerated by magnetospheric processes (e.g. substorms) are the cause of aurora and a source of ionization. The rate of electron energy loss and thus of ion production is a function of the particle energy and atmospheric density. More energetic particles penetrate deeper into the atmosphere producing the maximum ionization at lower altitudes. Ion production rates due to different electron energy are calculated in a model atmosphere (Rees, 1989) and the production rate must be integrated over energy of all precipitating particles. Although protons also produce aurora (proton aurora) a much more significant source of ionization are high energy protons emitted by solar flares. Enhanced fluxes of energetic protons are the cause of polar-cap absorption (PCA) events (Reid, 1974) that result in HF radio signal blanketing.

1.3.1.2   The Influence of Magnetic and Electric Fields on Plasma Convection

The geomagnetic poles are located at high geographic latitudes where the geomagnetic field axis is oriented vertically. The magnetospheric electric field **E** maps along the magnetic field lines down to the high-latitude ionosphere where **E** becomes basically horizontal. The mobility of charged particles is quite different in the directions perpendicular and parallel to the magnetic field **B**. Barring collisions, particles can move freely along the magnetic field, but can only move perpendicular to the magnetic field in the presence of electric fields. At high latitudes, the electric fields that are imposed on the ionosphere through interaction between the solar wind and magnetosphere drive horizontal motions of the ionospheric plasma due to $\mathbf{E} \times \mathbf{B}$ drift, so called ionospheric convection. The convection pattern mainly depends on the orientation of the IMF since the latter controls the solar wind coupling to the magnetosphere. In the simplest case, for purely southward IMF($B_z < 0$ and $B_y = 0$), the F-region plasma is convected antisunward across the polar cap and returns sunward along the dawn and dusk portions of the auroral oval. This two-cell convection pattern is tilted with respect to sun-earth line for IMF $B_y \neq 0$. The pattern is more complex for the northward IMF ($B_z > 0$) when the ionospheric convection is less intense and confined to higher latitudes (smaller polar cap) but more structured. The convection pattern may consist of three or four cells (Cowley and Lockwood, 1992). For the northward IMF, the magnetic reconnection between IMF and magnetosphere, which is thought to be the principal coupling process acting at the magnetopause, is significantly reduced and limited to tail lobe reconnection producing a circulatory motion of plasma

(lobe cell) in the central polar cap. In addition, viscous-like coupling process acting simultaneously at the magnetopause boundary may further complicate the convection morphologies. The ionospheric convection plays an important role not only in the plasma transport but also in structuring of the high-latitude ionospheric plasma (see, e.g., Crowley, 1996).

The motion of the plasma convecting at a velocity $\mathbf{v} = \mathbf{E} \times \mathbf{B}/B^2$ can be measured by a variety of techniques, including coherent- and incoherent-scatter radars, digisondes, satellites and rockets. The convection patterns can also be derived from equivalent currents based on ground magnetometer data. All of the techniques have their limitations but a network of HF radars is most suitable for large-scale mapping of the convection at high spatial and temporal resolution. Such capabilities have been successfully employed by Super Dual Auroral Radar Network (Greenwald et al., 1995). Statistical convection/potential patterns were first derived using the single radar data (Ruohoniemi and Greenwald, 1996) but many more details of convection dynamics are revealed from global imaging of the instantaneous convection using the full network of radars (Ruohoniemi and Baker, 1998). These observations are in general agreement with the conceptual model proposed by Cowley and Lockwood (1992).

1.3.1.3   Ionospheric and Field-Aligned Currents

Unlike the F region where the ionospheric plasma is to a first approximation collisionless (ion collision frequency v_i is much smaller than the ion gyrofrequency Ω_i) the ionosphere at E-region altitudes is a resistive medium. The ions are unmagnetized ($v_i \gg \Omega_i$) while the electrons are still "collisionless" ($v_e \ll \Omega_e$). At the E-region altitudes, the ion-neutral collisions are also important but the electron-neutral collisions are not. Thus, the electrons can move perpendicularly to magnetic field while the $\mathbf{E} \times \mathbf{B}$ drift of ions is impeded by collisions. This leads to an ionospheric (Hall) current. Generally, the dominant electric field is northward in the dusk sector and southward in the dawn sector of the auroral oval and the Hall conductivity (perpendicular to $\mathbf{E}$) dominates over the Pedersen conductivity (parallel to $\mathbf{E}$). Thus, the dominant electric current is an eastward electrojet in the afternoon and a westward electrojet in the morning. The ionospheric currents can be observed by ground magnetometers measuring the magnetic perturbations due to the horizontal electric currents, including the auroral electrojets.

The large value of the direct or longitudinal (along the magnetic field) conductivity suggests that significant currents should be able to flow along the magnetic field. In fact, field-aligned currents (FACs), sometimes called Birkeland currents after K. Birkeland who postulated their existence in 1908, were disputed by Chapman (1935) and promoted by Alfvén (1939) before the concept was finally confirmed in 1960s. The currents were first detected in the early 1970s by satellite-borne magnetometers that measured magnetic perturbations perpendicular to the field. Field-aligned currents provide an important electric circuit between the plasmas in the solar wind, magnetosphere and the high-latitude ionosphere. The system of large-scale field-aligned currents determined from low-altitude satellite measurements is actually a superposition of several current circuits connecting the ionosphere with different regions of the magnetosphere. The main FAC systems are composed of so-called Region 1 and Region 2, east-west aligned sheet currents that are connected in the ionosphere by the north-south Pedersen current.

1.3.1.4   Ionospheric Density Irregularities

In the E-region, strong auroral electrojets are sources of field-aligned ionospheric irregularities on perpendicular spatial scales from a fraction of 1 m to more than 100 m. If a differential electron-ion drift velocity $\mathbf{V}_d = \mathbf{V}_i - \mathbf{V}_e$ which is often taken approximately equal to the electron Hall drift $\mathbf{E} \times \mathbf{B}/B^2$, exceeds the ion-acoustic speed c_s (in the absence of horizontal electron density gradients) the plasma becomes unstable to Farley-Buneman (two-stream) instability. Electrostatic plasma density waves (fluctuations) are excited propagating nearly perpendicular to the magnetic field. When electron density gradients are present, two-stream and gradient drift theory is required to describe the instabilities (Farley and Fejer, 1975).

In the F-region, in the presence of horizontal electron density gradients, irregularities are produced by convective plasma processes and in particular, by the fluid $\mathbf{E} \times \mathbf{B}$ (gradient drift) interchange instability (Tsunoda, 1988). The irregularity scale-size spectrum extends from ~10 km down to the ion gyroradius. Because plasma density gradients are necessary for generating the F-region irregularities, a large-scale ionospheric structure is often of primary importance in the generation of small-scale structure. Much recent work has focussed on the modelling of plasma structuring in the F region using ever more sophisticated numerical codes (e.g., Gondarenko and Guzdar, 2004).

1.3.1.5   Large-Scale Ionospheric Structure

Large-scale plasma structure in the high-latitude ionosphere includes the auroral oval, the location of diffuse and discrete auroras, auroral and sun-aligned (transpolar) arcs, polar cap patches, boundary blobs, ionospheric troughs or plasma density depletions, and traveling ionospheric disturbances. The magnetospheric, ionospheric and atmospheric sources of this rather complex structure of the high-latitude ionosphere include solar induced ionization, soft and energetic particle precipitation, field-aligned and ionospheric currents, electric field enhancements, plasma convection, magneto-hydrodynamic waves (MHD) and atmospheric gravity waves (AGWs). However, the overarching process that controls, influences or modulates all of these sources is the solar wind coupling to the magnetosphere.

The most dynamic spatial and temporal changes of the ionospheric structure occur in the auroral zone, particularly during substorms. The optical manifestation of this structure is aurora borealis and aurora australis. The statistical distribution of aurora occurrence led to a concept of the auroral oval (Feldstein, 1963). The all-sky observations showed that at any given time the locus of aurora is not circular but oval, extending further equatorward in the midnight sector than on the dayside. Also, the auroral oval is widest near local midnight. Although originally proposed as a statistical concept, the auroral oval is present at all times, although its intensity varies with changing activity levels as can be seen from satellite images. The position and width of the auroral oval are both functions of the level of the solar wind-magnetosphere interaction. The aurora is a consequence of a complex interplay among energetic particle fluxes, electromagnetic waves that may alter the particle pitch angles, and the ionization potentials of the atmospheric atoms and molecules in the altitude range where the emissions originate. The conductivity changes that are caused by precipitating particles (mainly keV electrons) together with the convection electric field that maps from deep space into the auroral ionosphere, act together to regulate field-aligned currents and the ionospheric currents to which they connect. While most of the energy dissipated though Joule heating is associated with the large scale slowly varying Birkeland currents, Alfven waves with a Poynting flux into the ionosphere can also contribute significantly to ionospheric heating (Keilling et al., 2003). Non-visual techniques are now in place for routinely monitoring the location of auroral boundaries, e.g., Newell et al. (2002) and link to http://sd-www.jhuapl.edu/Aurora/ ovation_live/.

In the magnetospheric context, the auroral oval lies predominantly on closed field lines, with its poleward edge approximately marking the boundary between open and closed field lines. Energy is transported into the closed field line region from the magnetospheric boundary layers and through reconnection of magnetic field lines at a neutral line located beyond $\sim 20\ R_E$ in the magnetotail. Although the transport processes occur over huge volumes of space, regions where spatial gradients in pressure and electric field are present can feature parallel electric fields that accelerate electrons into the ionosphere in localized regions causing auroral arcs.

The nightside of the auroral oval can exhibit explosive auroral intensifications and rapid changes in arc structure, which are the mark of the auroral substorm. The auroral substorm is part of a global disturbance called a magnetospheric substorm, which features disturbances in all portions of the electromagnetic spectrum. A magnetospheric substorm features two distinctly different processes, namely directly driven and loading-unloading. Directly driven activity follows closely the increase and subsequent decrease in energy provided to the magnetosphere from the solar wind. Some of the energy entering the magnetosphere is stored

in the magnetotail and then is released at some later time. The auroral substorm is the visible signature of the release of that stored energy. Substorms last for 1-2 hours and may recur as often as every few hours; they are common when the IMF is southward and infrequent when the IMF is northward.

Ionospheric troughs are regions of depleted F-layer plasma density limited in latitude but extended in longitude that are observed in the sub-auroral and high-latitude ionosphere including the polar cap. Traditionally, a distinction is made between the mid-latitude (main) trough that is found adjacent and equatorward of the auroral oval and the high-latitude troughs that are observed inside the auroral oval or in the polar cap. Ionospheric convection plays a crucial role in trough formation through chemical and dynamical processes that include ion recombination, Joule heating and frictional heating. A subset of mid-latitude troughs are caused by rapid sub-auroral ion drift (SAID) due to large, short-lived electric fields occurring just equatorward of the auroral oval (Rodger et al., 1992).

The morphology of the polar ionosphere is largely determined by the orientation of the IMF which causes it to have one of two idealized characteristic states (Carlson, 1994). For southward IMF ($B_z < 0$), mesoscale (100-1000 km) plasma density enhancements known as polar patches are produced in the dayside ionosphere near the magnetospheric cusp footprint. The patches then convect antisunward in the polar cap where they can be observed by various techniques, principally radio and optical. Patches are associated with steep density gradients generating small-scale irregularities causing HF radar backscatter and severe scintillations of VHF and UHF radio signals. Although processes responsible for the formation of patches have been proposed a controversy still remains and several mechanisms are probably active in causing patches (Crowley, 1996). For northward IMF ($B_z > 0$), the polar ionosphere is characterized by considerable plasma convection structure and stable sun-aligned arcs. The sun-aligned or transpolar arcs delineate the polar cap plasma flow gradients/shears and provide valuable insight into magnetospheric topology and solar wind coupling processes (Carlson, 1994). It can be said that the polar ionosphere alternates between the two states depending on the IMF orientation but when IMF B_z switches sign often, both patches and sun-aligned arcs may coexist in the polar cap at any given time.

As already discussed in Section 1.2 for the mid-latitude ionosphere, the ionospheric F-region is often structured into wavelike fluctuations of electron density – travelling ionospheric disturbances (TIDs) – caused by atmospheric gravity waves (AGWs). While AGWs that originate in the lower atmosphere and propagate their energy upward may cause TIDs the bulk of large- and medium-scale TIDs are generated in the auroral oval by surges of auroral electrojet currents or convection electric field (Hocke and Schlegel, 1996). TIDs can be detected at all latitudes using various radio techniques including HF Doppler sounders, ionosondes, HF radars, incoherent scatter radars, or optically as airglow emissions modulated by AGWs. More recently TID-induced electron densities have been imaged using tomographic techniques (e.g., Pryse et al., 1995) and TEC measurements by GPS networks (e.g., Tsugawa et al., 2004). Modelling of gravity waves generated by enhancements in the ionospheric electric field shows that each electric field enhancement causes a Joule heating pulse, which in turn launches a gravity wave propagating equatorward and poleward of the source region (Millward et al., 1993a, b). To elucidate the physics of the AGW-TID relationship, theoretical simulations were conducted and compared with incoherent scatter radar data (Kirchengast 1996; Kirchengast et al., 1996). This comparison allowed comprehensive information on AGWs and the AGW-induced TID ionospheric parameters (electron density, field-aligned ion drift, and ion and electron temperatures), or rather their variations with respect to background values based on a model of ionosphere, to be deduced by means of careful physical modelling.

The following sections focus on specific phenomena/processes and their effects on ionosphere and its structure in more detail. Considerable efforts that have been expended to understand the morphology of the ionosphere as well as magnetospheric processes have produced a vast scientific literature. Appropriate references for further details are made wherever deemed necessary but this chapter does not claim to be an exhaustive review. Rather it should be considered as a concise summary to provide an interested reader with up-to-date information (or a source of it) on the morphology of high-latitude ionosphere.

1.3.1.6   References

Alfvén, H., A theory of magnetic storms and of the aurorae, *K. Sven. Vetenskapakad.*, Handl., Ser. 3, 18(3), 1939. (Reprinted in part with comments by Alex Dessler and John Wilcox in *Eos*, 51, 180-194, 1970).

Carlson, H. C., Jr., Dark polar ionosphere: Progress and future challenges, *Radio Sci.*, 29(1), 157-165, 1994.

Chapman, S., The electric currents system of magnetic storms, *Terr. Mag. Atmos.*, 40, 349-370, 1935.

Cowley, S. W. H., and M. Lockwood, Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system, *Ann. Geophys.*, 10, 103-115, 1992.

Crowley, G., Critical review of ionospheric patches and blobs, *Review of Radio Science 1993-1996*, 619-648, Oxford Univ. Press, New York, 1996.

Farley, D. T., and B. G. Fejer, The effect of the gradient drift term on Type 1 irregularities, *J. Geophys. Res.,* 80, 3087-3090, 1975.

Feldstein, Y. I., On Morphology and Auroral and Magnetic Disturbances at High Latitudes, *Geomagn. Aeron.* 3, 138, 1963.

Gondarenko, N.A., and P.N. Guzdar, Plasma patch structuring by the nonlinear evolution of the gradient drift instability in the high-latitude ionosphere, *J. Geophys. Res.*, 109, A09301, doi:10.1029/2004JA010504, 2004.

Greenwald, R.A., K.B. Baker, J.R. Dudeney, M. Pinnock, T.B. Jones, E.C. Thomas, J.-C. Villain, J.-C. Cerrisier, C. Senior, C. Hanuise, R.D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A.D.M. Walker, N. Sato, and H. Yamagishi, DARN/SUPERDARN A global view of the dynamics of high-latitude convection, *Space Sci. Rev.*, 71, 761 796, 1995.

Hocke, K., and K. Schlegel, A review of atmospheric gravity waves and traveling ionospheric disturbances: 1982-1995, *Ann. Geophysicae*, 14, 917-940, 1996.

Keiling, A., J. R. Wygant, C. A. Cattell, F. S. Mozer, and C. T. Russell, The Global Morphology of Wave Poynting Flux: Powering the Aurora, *Science*, 299, 383-386, 2003.

Kirchengast, G., Elucidation of the physics of the gravity wave-TID relationship with the aid of theoretical simulations, *J. Geophys. Res.*, 101, 13353-13368, 1996.

Kirchengast, G., K. Hocke, and K. Schlegel, The gravity wave-TID relationship: insight via theoretical model - EISCAT data comparison, *J. Atmos. Terr. Phys.*, 58, 233-243, 1996.

Millward, G.H., S. Quegan, R.J. Moffett, T.J. Fuller-Rowell, and D. Rees, A modeling study of the coupled ionospheric and thermospheric response to an enhanced high-latitude electric field event, *Planet. Space Sci.*, 41, 45-56, 1993a.

Millward, G.H., R.J. Moffett, S. Quegan, and T.J. Fuller-Rowell, Effect of an atmospheric gravity wave on the mid-latitude ionospheric F layer, *J. Geophys. Res.*, 98, 19173-19179, 1993b.

Newell, P.T., T. Sotirelis, J.M. Ruohoniemi, J.F. Carbary, K. Liou, J.P. Skura, C.-I. Meng, C. Deehr, D. Wilkinson, and F.J. Rich, OVATION: Oval variation, assessment, tracking, intensity, and online nowcasting, *Annales Geophysicae*, 20, 1039-1047, 2002.

Pryse, S. E., C. N. Mitchell, J. A. T. Heaton and L. Kersley, Travelling ionospheric disturbances imaged by tomographic techniques, *Ann. Geophysicae*, **13**, 1325-1330, 1995.

Rees, M. H., *Physics and Chemistry of the Upper Atmosphere*, Cambridge Univ. Press, New York, 1989.

Reid, G.C. 1974. Polar cap absorption - Observations and theory. *Fundamentals of Cosmic Physics* 1:167-202.

Rodger, A.S., R.J. Moffett, and S. Quegan, The role of ion drift in the formation of ionisation troughs in the mid- and high-latitude ionosphere – a review, *J. Atmos. Terr. Phys*., 54, 1-30, 1992.

Ruohoniemi, J. M. and R. A. Greenwald, Statistical patterns of high-latitude convection obtained from Goose Bay HF radar observations, *J. Geophys. Res.*, 101, 21743, 1996.

Ruohoniemi, J. M. and K. B., Baker, Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations, *J. Geophys. Res.*, 103, 20797-20811, 1998.

Tsugawa, T., A. Saito, and Y. Otsuka, A statistical study of large-scale traveling ionospheric disturbances using the GPS network in Japan, *J. Geophys. Res.*, 109(A6), A06302, doi:10.1029/2003JA010302, 2004.

Tsunoda, R. T., High latitude F-region irregularities: A review and synthesis, *Rev. Geophys., 26,* 719, 1988.

1.3.2   Particle Precipitation

(Dirk Lummerzheim, Geophysical Institute, University of Alaska)

1.3.2.1   Introduction

The most spectacular manifestation of particle precipitation into the thermosphere is the aurora. Energetic electrons in the energy range of a fraction of a keV to several keV are the cause of the visible aurora, but energetic proton precipitation also causes diffuse auroral emissions. The electron precipitation averaged over the entire auroral oval dominates the auroral precipitation. The spatial distribution of protons and electrons is, on average, different, and the equatorward edge of the evening aurora tends to be dominated by proton precipitation. During large substorms, the hemispheric power in the auroral electron precipitation can reach several hundred GW. Locally, energy fluxes in small scale auroral filaments are observed to reach up to several hundred mW/m^2. These precipitating particles deposit their energy and cause ionization at altitudes between 80 and 500 km. The source of the auroral particles is in the magnetosphere. In the inner magnetosphere pitch angle scattering produces electrons with high mean energy (tens of keV) and low energy fluxes, which in turn give rise to the diffuse aurora at the equatorward boundary of the auroral oval. The plasma sheet and tail of the magnetosphere is the region where substorms originate. On the dayside in the cusp, the auroral precipitation originates at the magnetopause. While nightside (substorm) precipitation typically has mean energies of several keV to more than 10 keV for electrons and several tens of keV for protons, the precipitation in the cusp is softer and causes higher altitude aurora with dominant OI(630 nm) emissions from the F-region.

Non-auroral precipitation occurs inside the polar cap and in regions that are magnetically connected to the sunlit part of the other hemisphere. Photoelectrons that are generated inside the thermosphere escape upwards and travel along the magnetic field to the other hemisphere. Outside of the auroral region and on the nightside, these fluxes can be the dominant precipitation, although the energy fluxes are very low (below 1 mW/m^2) with mean energies of a fraction of a keV.

The polar caps are magnetically connected to the magnetosheath and the solar wind has direct access to the polar cap ionosphere. Electrons of a few hundred eV that originate in the solar corona precipitate here. Figure 1-4 shows an example of the auroral electron and ion precipitation as well the soft electrons in the polar cap. During large solar events, high energy protons (10 MeV) also penetrate into the lower ionosphere at high latitudes.
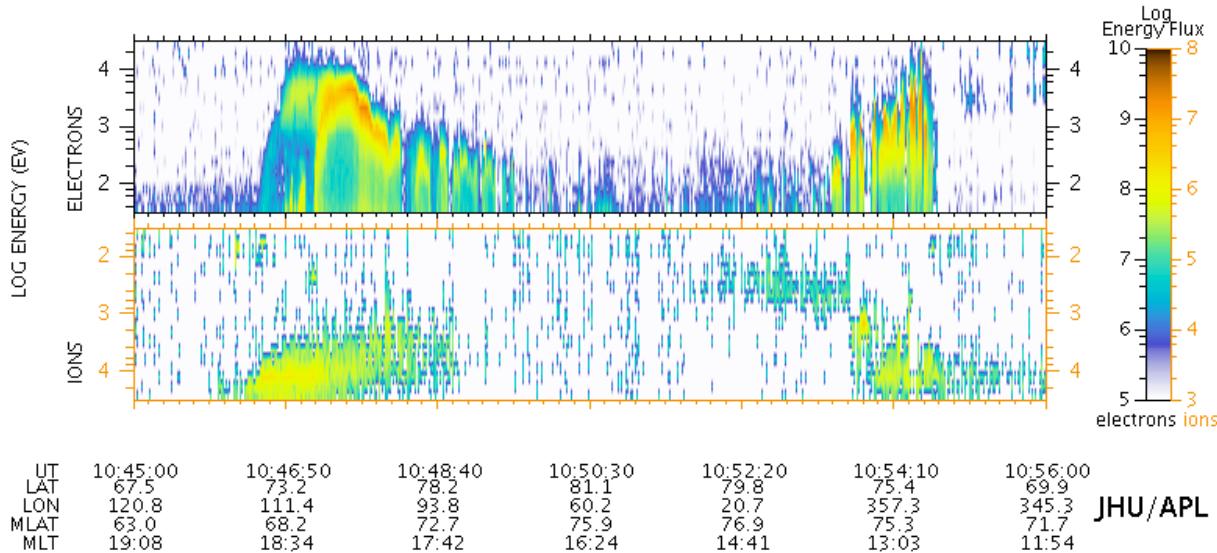
**Figure 1-4: Example of the Downward Electron and Ion Flux from
a DMSP Pass Over the Northern Auroral Oval.**

The satellite enters the aurora on the evening side and observes electrons of several keV energy. As it enters the polar cap, low fluxes in the range of about 100 eV are seen, until eventually the satellite exits the auroral oval near local noon (13:00 MLT). Close inspection of the ion fluxes show that in the evening sector the ion precipitation extends to latitudes below the electron precipitation.

1.3.2.2   Auroral Electron Precipitation

1.3.2.2.1   Modeling

Auroral electrons are guided by the Earth's magnetic field into the thermosphere. For the field strength and average mean free path at auroral altitudes the gyrofrequency is much larger than the collision frequency. This allows specifying the motion of an individual electron by the motion of its guiding center. Consequently, the motion is symmetric with respect to azimuth and one-dimensional along the direction of the magnetic field. The acceleration of the electrons takes place in the magnetosphere, while electrons only lose energy as they penetrate the collision dominated thermosphere. The time that it takes the electron to penetrate the thermosphere and deposit its energy is small compared to the time it takes for the host medium to change in response to the precipitating electrons. The collision processes are dominated by the electron neutral collisions, and the non-linearity that would arise from collisions with aurorally produced ions are negligible. All collisions are assumed to be binary collisions, i.e., no three-body collisions need to be considered. The density of streaming particles is low compared to the ambient density. The mirror force has a small effect on the precipitating electrons and is also usually neglected. These assumptions allow to simplify the non-linear Boltzmann equation for auroral electrons and to derive a linear, steady state transport equation:

$$\mu \frac{\partial I(E,\mu,\tau)}{\partial \tau} = I(E,\mu,\tau) - \frac{\omega}{2} \int_{-1}^{1} d\mu' p(\mu,\mu') I(E,\mu',\tau) - Q\left(I(E'>E),\mu,\tau\right)$$

In this transport equation, the change of the electron intensity (or differential number flux) I as a function of energy E, cosine of the pitch angle μ, and scattering depth τ, as it penetrates deeper into the atmosphere is due to loss from collisions (first term on the right), angular redistribution from elastic scattering (second term) and energy degradation and secondary electron production Q (third term). The angular redistribution in the elastic scattering term is given by the phase function p, while the scattering albedo ω is given by the ratio of the total elastic cross-section σ^{el} to the total cross section σ^{tot}:

$$\omega = \frac{\sum n(z)\sigma^{el}(E)}{\sum n(z)\sigma^{tot}(E)}$$

where $n(z)$ is the neutral density as function of altitude and the sums extend over all neutral species. The scattering depth is defined as:

$$\tau(z,E) = \sum \sigma^{tot}(E) \int_{z}^{\infty} n(z')\, dz'$$

Because of the assumption that electrons do not gain energy once they enter the thermosphere, the energy redistribution term can be expressed in terms of the electron intensity at higher energies, and the transport equation can be solved starting at some high energy and progressing to lower energies. The numerical solution requires that neutral densities, cross sections for various collision processes, phase functions, energy redistribution functions, and secondary electron production from ionization processes are specified. Lab measurements of many of these processes are available but often have large uncertainties. Neutral density models that are based on global observations such as MSIS [*Hedin*, 1991, 1992; *Picone et al.*, 2002] also have a poor representation of the thermosphere in the auroral oval and introduce uncertainties into an auroral model.

Two complications in a numerical treatment of the transport equation arise from the range of densities, and thus the scattering depth for the altitude region in which the equation must be solved, and from the large difference between the mean energy of the incident electrons and the energy loss in individual collisions. Either the numerical grid has to resolve these large ranges of the required variable space or numerical methods have to be found to accurately treat these mismatched variables.

A number of different approaches to the solution of the transport equation are presented in the literature. Some models relax the assumptions listed above, usually with a penalty of introducing other assumptions. Electron transport models for Earth's aurora have been developed by *Banks and Nagy*, [1970; see also *Nagy and Banks*, 1970] using a two-stream approach, *Banks et al.* [1974] using Monte-Carlo techniques, *Jasperse* [1976] using a theoretical method of solution, *Strickland et al.* [1976] using a multi-stream approach (i.e., with pitch angle resolution), *Stamnes* [1980, 1981] using a two-stream discrete ordinate method, *Link* [1992] applying a Feautrier solution, *Lummerzheim and Lilensten* [1994] using a discrete ordinate technique, and *Solomon* [1993] using a Monte-Carlo technique. *Min et al.* [1993] took the discrete ordinate method of *Lummerzheim and Lilensten* [1994] to include small field aligned ionospheric electric fields in order to study the influence of the ambipolar diffusion field on electron precipitation. *Peticolas and Lummerzheim* [2000] have developed a time-dependent electron transport model which can simulate flickering aurora or fast moving auroral filaments.

The solution of the transport equation provides the electron intensity as a function of altitude, given an electron source. Figure 1-5 shows an example of the downward electron flux from a model calculation.

For photoelectrons this source would be distributed in altitude and must be calculated from a radiative transfer calculation of solar EUV radiation, while for aurora the incident electron precipitation can be specified as a downward directed electron flux at the upper boundary of the model. From the altitude profile of the electron intensities, one can calculate the excitation and ionization rates, energy deposition and heating rates, electron particle and energy fluxes, and a number of auroral emission rates.
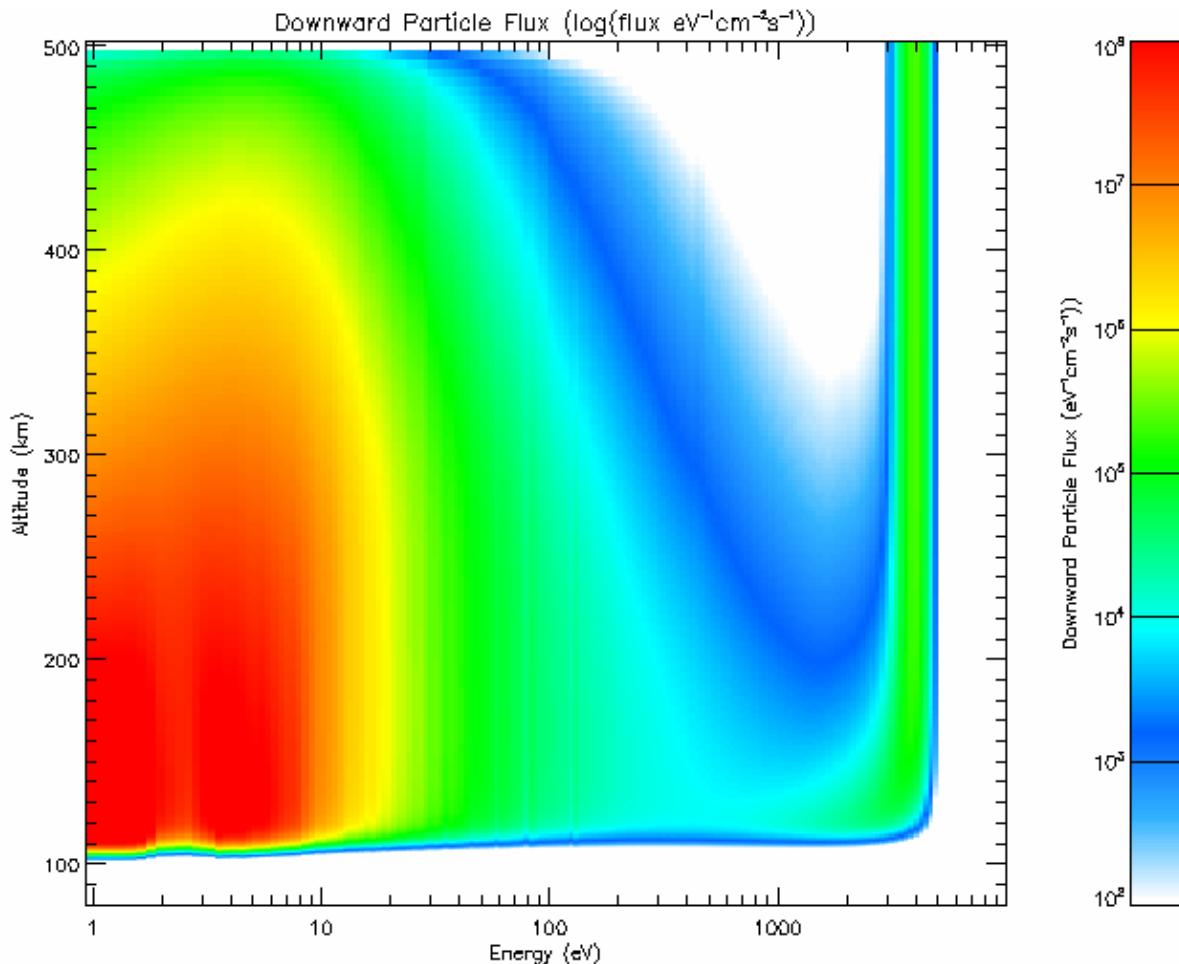
**Figure 1-5: Example of the Downward Electron Flux as a Function
of Altitude and Energy from a Model Calculation.**

The incident flux was specified at 500 km as a Gaussian energy distribution at 4 keV with an energy flux of 1 mWm^{-2}. As the electron penetrate the atmosphere, low energy electrons from ionization collisions are generated. The high energy flux diminishes through excitation and ionization collisions as can be seen at about 120-150 km altitude. Auroral emissions from excited states are mostly generated by collisions of the secondary electrons in the energy range of a few eV to several tens of eV, where cross sections maximize and the secondary electron fluxes are large.

1.3.2.2.2   Observations

The spectrum of the precipitating electrons can be obtained either by direct in situ observations from sounding rockets or satellites, or by inferring from optical auroral or ionospheric radar observations. In situ observations are the least ambiguous, while remote sensing using optical observations and modeling can provide simultaneous large scale maps of auroral particle precipitation characteristics.

Low earth orbit satellites typically move at velocities of 7-10 km/s, and the in situ observations have to be made at a high frequency in order to resolve the spatial structure of auroral precipitation. An example of the extreme fine structure is presented by *Boehm et al.* [1995]. The Freja satellite had a high resolution electron particle spectrometer which obtained the energy spectrum by sweeping the high voltage in the detector. During a transition through an auroral precipitation region, the instrument encountered a rapid variation in the observed electron flux as a function of energy which *Boehm et al.* interpreted as a sharp spatial rather than temporal gradient that the satellite penetrated while sweeping though the energy channels of the detector. The 2 ms rise time translates into a spatial length scale of 10 m for the auroral electron precipitation.
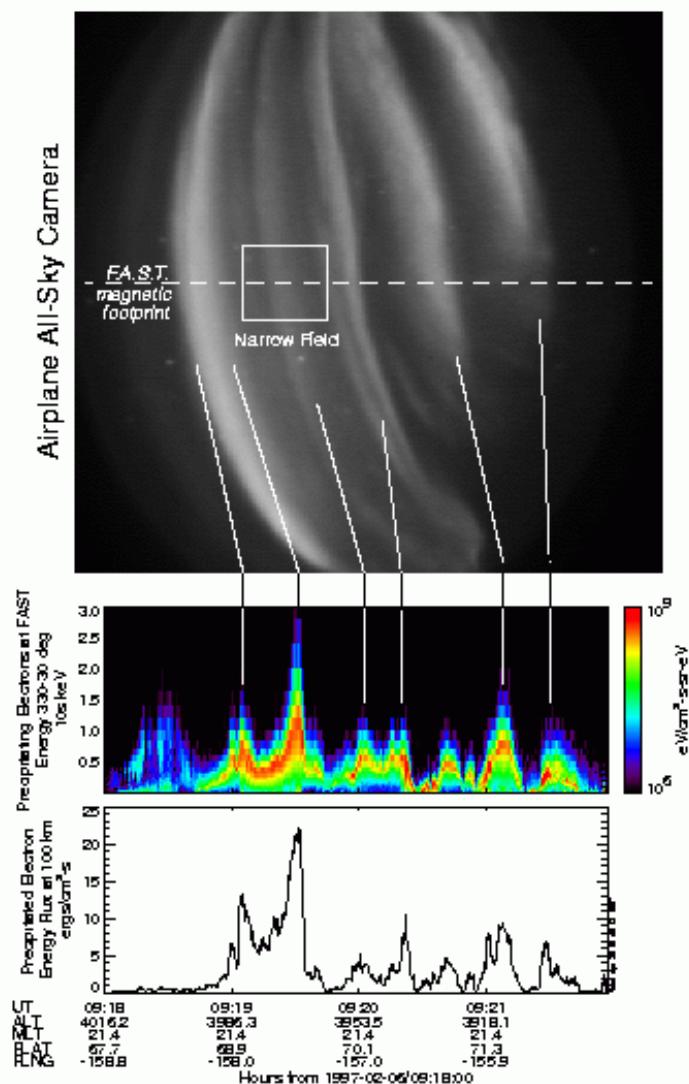
Figure 1-6: Electron Precipitation from FAST Observations with Coincident Auroral Imaging from an Airplane – from Stenbaek-Nielsen et al., 1998.

Such high resolution measurements are not made routinely, and most in situ observations are therefore averages over a limited spatial region. The DMSP satellites carry particle detectors that observe the incident electron and proton spectra in the vertical direction and collect one spectrum over an integration period of 1 s [*Hardy et al.,* 1985]. These electron precipitation spectra have been compiled into statistical

descriptions of the auroral oval [*Hardy et al., 1985*], have been used to identify precipitation boundaries [*Newell and Meng, 1988; Newell at al., 1991*], and have provided the electron precipitation in numerous case studies of aurora. A statistical analysis of the DMSP obtained particle spectra lead *Newell et al.* [1996] to the conclusion that the mean energy of auroral electrons in sunlit aurora is on average less than in aurora in the dark thermosphere. These findings are supported by subsequent publications of the analysis of UVI auroral images [*Liou et al.*, 1997], and the re-analysis of an old data set of ground based observations [*Deehr et al., 2005*].

Most particle detectors for auroral precipitation measurements cover the energy range from a few eV to about 32 keV. This is the energy range where most of the energy flux is and which contains the mean energy of the precipitation. The UARS PEM (particle environment monitor) and the NOAA/TIROS satellites also carry integral detectors to extend the energy sensitivity into the MeV range. Although the electrons in the hundreds of keV and MeV range contribute little to the auroral emissions, these electrons penetrate deeper into the middle atmosphere where they may be the dominant source for excitation, ionization, and energy deposition [*Winningham et al., 1993, Codrescu et al.,* 1997]. Ground based cosmic noise absorption measurements are suitable to determine particle fluxes that penetrate into the D-region [*Tanaka et al.*, 2005].

The FAST satellite was designed to obtain high resolution observations of auroral particle precipitation and other auroral parameters [*Carlson et al.,* 1998]. The particle detector obtains a full spectrum (pitch angle and energy distribution) every 70 ms. A high time resolution mode allows a limited number of energies and pitch angle measurements with integration times of 1.7 ms. A review of the large number of publications and science results from the FAST satellite exceeds this article by far. There are numerous other satellites and sounding rocket observations of precipitating particles and their characteristics.

Indirect methods to obtain the auroral electron precipitation are based on spectroscopic optical observations, or incoherent scatter radar observations. Incoherent scatter radar measures the altitude profile of the plasma density in the ionosphere. In aurora, this can be related to the ionization rate profile with some assumptions and using an auroral model one can deduce some parameters describing the incident electron flux [*Robinson and Vondrak, 1985, Doe at al., 1997*]. *Lanchester et al.* [1994, 1997] used high resolution (200 ms temporal resolution) EISCAT measurements of the plasma density profile and an ionospheric model to conclude that small auroral filaments are characterized by near mono-energetic electron precipitation of several keV mean energy and energy fluxes in excess of 100 mWm^{-2}.

Optical methods to obtain the mean energy of the auroral electrons are based on the fact that the altitude of maximum emission is a function of the mean energy. The neutral composition changes with altitude, and ratios of emissions can therefore be used to measure this altitude and thus infer the mean energy. Because of the lighter mass and greater scale height of the atomic oxygen, the ratios of molecular nitrogen (or molecular nitrogen ion) emissions and atomic oxygen emissions are most useful. The disadvantage of this method is that a good model of the neutral atmosphere is required. The aurora modifies the ionosphere and subsequently the neutral composition sufficiently, that the atomic oxygen density in aurora is not very well represented in statistical neutral atmosphere models like MSIS. *Hecht et al.* [1989] has developed techniques where the comparison of three different auroral spectral brightnesses can be used to obtain the auroral electron energy and a modification of the atomic oxygen density profile at the same time.

Remote sensing from satellite in the UV wavelength range has the advantage that the auroral brightness can be analyses on the entire auroral oval, including the sunlit portion [*Lummerzheim et al., 1997*]. The altitude of the auroral emission maximum can then be inferred from the N_2LBH spectral brightness. The LBH bands cover a sufficient wavelength range that the short wavelength end of the bands are strongly absorbed by O_2 Schuman-Runge extinction, while the long wavelength bands propagate nearly without extinction. This is used in the analysis of Dynamics Explorer images [*Rees et al, 1988*], Polar UVI images [*Germany et al, 1998*], and TIMED GUVI [*Strickland et al., 1992*].

1.3.2.3   Auroral Ion Precipitation

1.3.2.3.1   Modeling

In order to model proton aurora, the assumptions made to derive an electron transport equation lead to a similar equation for protons. There is, however, one significant complication. Protons easily charge-exchange in collisions with atomic oxygen, thus turning into hydrogen. The newly created hydrogen atom keeps nearly the same kinetic energy as the original proton had, but does not gyrate around the magnetic field. The precipitation therefore spreads horizontally as soon as collisions with the neutrals in the thermosphere occur. The energetic hydrogen can be stripped of its electron in subsequent collisions. This process can repeat several times before the particle has deposited its energy. The energetic hydrogen is described by the same transport equation as the proton, except that the charge-exchange collision term is replaced by the stripping collision term. Ionization collisions as well as stripping collisions create electrons which have energies up to several hundred eV. These electrons must be included in a proton aurora model by solving an electron transport equation with a source term to represent this secondary electron production.

Because the proton and hydrogen transport equations are strongly coupled, the numerical solution of these equations offers additional challenges over the modeling of electron aurora. The spreading across the magnetic field direction makes this an inherently two-dimensional problem. However, there are few two or three dimensional numerical simulations available. Most authors make the additional assumption that the precipitation region is large and can be treated as a one-dimensional problem. To account for the finite extend of incident proton beams, a spreading factor which is either derived from observations of proton aurora or from parameterizations of three-dimensional calculations is used.

One of the first numerical techniques to solve these transport equations was the Monte-Carlo method (*Davidson, 1965)*. Davidson's model included the coupled proton-hydrogen transport in two dimensions to assess the spreading of the incident proton precipitation. More recent Monte-Carlo simulations with different emphasis were performed by *Kozelov* [1993] who developed a proton-hydrogen auroral model which includes a converging magnetic field, *Decker et al.* [1996] who compared a number of different approaches to solve the coupled transport equations and used the Monte-Carlo simulation as a validation tool, *Lorentzen et al.* [1998] looked in particular at hydrogen emissions in the cusp, *Synnes et al.* [1998] developed a Monte-Carlo simulation including the angular redistribution to study the proton aurora as a source for ENA (energetic neutral atoms) observed from space, and *Gérard et al.* [2000] who developed the proton auroral model for the interpretation of space based auroral images. *Jasperse and Basu* [1982] and *Basu et al.* [1987] constructed quasi-analytical solutions for the coupled one-dimensional transport equations. Explicit numerical solutions to the coupled transport equations for proton and hydrogen precipitation, and the additional coupled equation for the secondary electrons were published by *Basu et al.* [1993], *Strickland et al.* [1993] and *Galand et al.* [1997, 1998].

Hydrogen emissions are negligible in electron aurora because the ambient hydrogen density is too low to lead to any significant emission. Proton aurora produces its own hydrogen, and these emissions are characteristic for proton precipitation. The hydrogen that is produced by charge exchange collisions has the kinetic energy of the proton precipitation. The hydrogen emissions have therefore a distinctive Doppler shifted emission line profile. As the proton-hydrogen beam forms in the atmosphere, energy degradation and scattering will produce a smooth energy distribution of particles even from a mono-energetic incident beam. This energy distribution is reflected in the hydrogen emission line profile. Scattering processes and the convergent magnetic field also produce angular redistribution of the incident beam. Part of the energetic hydrogen is then moving upwards, giving rise to a Doppler shift in the opposite direction. For a ground based observer, the bulk of the emissions come from downward moving particles, thus creating a blue shifted line profile. Angular distribution generates upward moving particles which cause a red-shifted emission. *Gérard et al.* [2000] make use of this Doppler shift to separate the auroral Lyman-α emissions from the geocorona emissions which are not Doppler shifted. The IMAGE satellite makes use of this

method to obtain global images of the proton aurora [*Mende et al.*, 2000]. *Galand et al.* [1997, 1998] have studied the angular distribution in detail with their multi-stream proton aurora model. High spectral resolution ground based observations of the Doppler shifted H_β line show the expected blue shifted line profile and the small red-shifted component [*Lummerzheim and Galand,* 2001]. Figure 1-7 demonstrates this by comparing an auroral H_β line profile with a stationary calibration lamp.
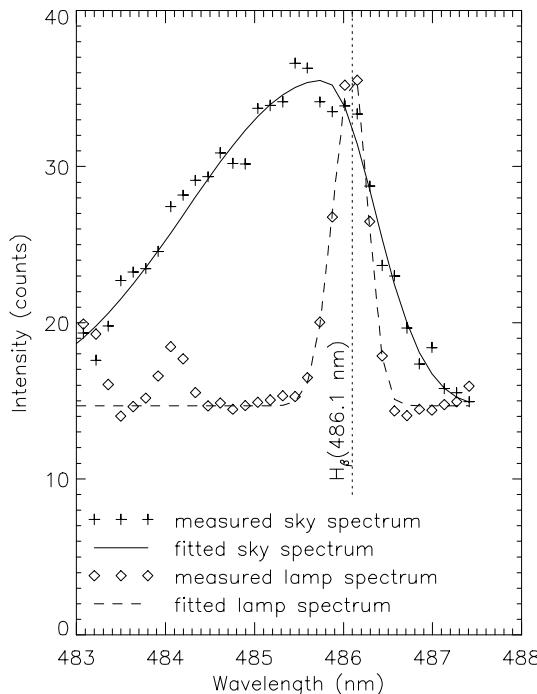
Figure 1-7: High Spectral Resolution Observations of the Doppler H_β Line Profile from Proton Aurora. Superimposed is the unshifted line shape from a stationary hydrogen lamp, taken with the same instrument and resolution.

The proton-hydrogen stream and the secondary electrons also cause excitation of the ambient neutral constituents of the thermosphere in aurora and give rise to all the emissions that are seen in electron aurora. Protons are unlikely to produce excitation without ionization, but the energetic hydrogen can cause excitation of all neutral species. The secondary electrons have a lower mean energy than secondary electrons in electron aurora and can thus cause slightly different brightness ratios between auroral emissions [Lummerzheim et al., 2001]. This can cause erroneous interpretation of auroral brightness ratios in terms of mean energy, if the brightness is assumed to be due to electron precipitation [Galand and *Lummerzheim,* 2004].

1.3.2.3.2 *Observations*

Most of the satellites and sounding rockets that carry electron detectors also have detectors for energetic ions. These may include detectors for measuring the ion mass and pitch angle distribution in addition to the energy spectrum as on FAST [*Carlson et al.,* 1998], or they may only determine the energy flux and mean energy from onboard processing of a measured energy spectrum (e.g. NOAA-TIROS Total Energy Detector (TED) [*Evans*, 1987]). Like the electron spectra, statistical analysis of a large number of satellite passes have been conducted [e.g. *Hardy et al.,* 1985; *Newell and Meng,* 1988; *Newell at al.,* 1991]. Case studies where in situ proton precipitation, satellite Lyman-α images, and ground based high resolution spectroscopy of the Doppler shifted H_β line have been use to validate model calculations [*Lanchester et al.*, 2003] rely on measured proton precipitation energy spectra as model input.

Although the majority of the auroral ion precipitation is in the form of protons, great geomagnetic storms have also caused energetic oxygen ion and neutral precipitation [*Stephan et al.*, 2004], usually at subauroral latitudes. The source of these oxygen atoms is in the ring current. This oxygen is observed by enhancements of atmospheric oxygen emissions and by ENA (energetic neutral atom) imaging from space [*Stephan et al.*, 2000].

1.3.2.4   Polar Cap

Particle precipitation also occurs outside the auroral oval. The polar cap is magnetically connected to the solar wind, and solar wind electrons and ions have direct access to the ionosphere. This precipitation is called polar rain [Winningham and Heikkila, 1974]. It is characterized by very low energy fluxes (compared to auroral precipitation) of typically 10^{-3} mWm^{-2} and a mean energy typically in the range of a few hundred eV. On rare occasions, the electron flux can be more energetic. Newell and Meng [1990] conducted a study over three years and found 17 days where electron flux exceeded a threshold of 10^6 cm^{-2}s^{-1} at an energy above 1 keV. Intense polar rain events are not associated with geomagnetic storms, but rather occur during extreme quiet times [Hardy, 1984]. One the "day when the solar wind vanished" (11 May, 1999) intense polar rain was observed with energies up to 10 keV, causing X-ray emissions at extremely high latitude [Anderson et al., 2000]. Because of the usually low energy, the ionospheric effects of polar rain are mostly in the F-region. The polar rain can cause high airglow emissions of OI(630 nm) inside the polar cap [McEwen and Harrington, 1991].

On the other extreme of particle precipitation are the very high energy protons during Solar Proton events (SPE). These protons have energies in excess of 10 MeV, and have fluxes of more than 10 particles cm^{-2} s^{-1} ster^{-1}. SPE are rare, however, the Space Environment Center (SEC) has recorded 134 such events in the 31-year period from 1976-1997 [Kunches and Zwickl, 1999]. Since these protons penetrate into the lower ionosphere they have the potential to disrupt HF radio communication. Because of the large gyroradius of the high energy protons, the geomagnetic field limits access of this precipitation in latitude as a function of the rigidity. Ground based global riometer observations can thus be used to determine some energy characteristics based on the latitudinal extend of the D-region ionization [Rodger et al., 2006]. Depending on their energy, these energetic protons also penetrate into the middle atmosphere where they can have a large influence on the local chemistry, including the ozone layer [*Jackman et al.*, 2004]. These very high energy protons are also a health risk for astronauts [e.g. *Hoff et al.*, 2004].

1.3.2.5   References

Anderson, P. C., D. L. McKenzie, D. W. Datlowe, J. D. Hawley, S. M. Petrinec, M. Schulz, and D. E. Larson, 2000, Polar cap x-ray and electrons under low density solar wind conditions: Coordinated PIXIE and DMSP observations on 11 May 1999, *Geophys. Res. Lett., 27,* 4021.

Banks, P. M., and A. F. Nagy, 1970, Concerning the influence of elastic scattering upon photoelectron transport and escape, *J. Geophys. Res.*, *75*, 1902.

Banks, P. M., C. R. Chappell, and A. F. Nagy, 1974, A new model for the interaction of auroral electrons with the atmosphere: spectral degradation, backscatter, optical emission, and ionization, *J. Geophys. Res.*, *79*, 1459.

Basu, B., J. R. Jasperse, R. M. Robinson, R. R. Vondrak, and D. S. Evans, 1987, Linear transport theory of auroral proton precipitation: A comparison with observations, *J. Geophys. Res.*, *92*, 5920.

Basu, B., J. R. Jasperse, D. J. Strickland, and R. E. Daniell Jr., 1993, Transport-theoretic model for the electron-proton-hydrogenatom aurora, 1, Theory, *J. Geophys. Res.*, *98*, 21,517.

Carlson, C. W., R. F. Pfaff, and J. G. Watzin, 1998, The Fast Auroral SnapshoT (FAST) mission, *Geophys. Res. Lett., 25*, 2013.

Codrescu, M. V., T. J. Fuller-Rowell, R. G. Roble, and D. S. Evans, 1997, Medium energy particle precipitation influences on the mesosphere and lower thermosphere, *J. Geophys. Res., 102*, 19, 977.

Davidson, G. T., 1964, Expected spatial distribution of low energy protons precipitated in the auroral zones, *J. Geophys. Res., 70*, 1061.

Deehr, C. S., M. H. Rees, A. E. H. Belon, G. J. Romick, and D. Lummerzheim, 2005, Influence of the ionosphere on the altitude of discrete auroral arcs, *Ann. Geophys., 23,* 759.

Doe, R. A., J. D. Kelly, D. Lummerzheim, M. Brittnacher, G. A. Germany, G. K. Parks, and J. F. Spann, 1997, Initial comparison of POLAR UVI and Sondrestrøm IS radar estimates for auroral electron energy flux, *Geophys. Res, Lett., 24*, 999.

Evans, D. S., 1987, Global Statistical Patterns of Auroral Phenomena, in *Proceedings of the Symposium on Quantitative Modeling of Magnetospheric-Ionospheric Coupling Processes*, 325, Kyoto.

Galand, M., J. Lilensten, W. Kofman, and R. B. Sidje, 1997, Proton transport model in the ionosphere, 1, Multistream approach of the transport equations, *J. Geophys. Res., 102*, 22, 261.

Galand, M., J. Lilensten, W. Kofman, and D. Lummerzheim, 1998, Proton transport model in the ionosphere, 2, Influence of magnetic mirroring and collisions on the angular redistribution in a proton beam, *Ann. Geophys., 16*, 1308.

Galand, M., and D. Lummerzheim, 2004, Contribution of proton precipitation to space-based auroral FUV observations, *J. Geophys. Res., 109*, A03307 10.1029/2003JA010321.

Gérard, J.-C., B. Hubert, D. V. Bisikalo, and V. I. Shematovich, 2000, A model of the Lyman-α line profile in the proton aurora, *J. Geophys. Res., 105*, 15, 795.

Germany, G. A., G. K. Parks, M. J. Brittnacher, J. F. Spann, J. Cumnock, D. Lummerzheim, F. Rich, and P. G. Richards, 1998, Energy characterization of a dynamic auroral event using GGS UVI images, AGU Monograph 104 "Encounter Between Global Observations and Models in the ISTP Era", Jim Horwitz, Dennis Gallagher, and Bill Peterson, editors, *143*.

Hardy, D. A, 1984, Intense fluxes of low-energy electrons at geomagnetic latitudes above 85 deg, *J. Geophys. Res., 89,* 3883.

Hardy, D. A., M. S. Gussenhofen, and E. Holman, 1985, A statistical model of auroral electron precipitation, *J. Geophys. Res., 90*, 4229.

Hecht, J. H., A. B. Christensen, D. J. Strickland, and R. R. Meier, 1989, Deducing composition and incident electron spectra from ground-based auroral optical measurements: Variations in oxygen density, *J. Geophys. Res., 94*, 13,553.

Hedin, A. E., 1991, Extension of the MSIS thermosphere model into the middle and lower atmosphere, *J. Geophys. Res., 96*, 1159.

Hedin, A. E., 1992, MSIS model (1986), *Planet. Space Sci., 40*, 555.

Hoff, J. L., L. W. Townsend, and E. N. Zapp, 2004, Interplanetary crew doses and dose equivalents: variations among different bone marrow and skin sites, *Adv. Space Res., 34,* 1347.

Jackman, C. H., M. T. Deland, G. J. Gordon, E. L. Fleming, D. K. Weisenstein, M. K. W. Ko, M. Sinnhuber, J. Anderson, and J. M. Russell, The influence of the several very large solar proton events in years 2000 2003 on the neutral middle atmosphere, *Adv. Space Res., 35,* 445.

Jasperse, J. R., 1976, Boltzmann-Fokker-Planck model for the electron distribution function in the earth's atmosphere, *Planet. Space Sci.*, *24*, 33.

Jasperse, J. R., and B. Basu, 1982, Transport theoretic solutions for auroral proton and H atom fluxes and related quantities, *J. Geophys. Res*., *87*, 811.

Kozelov, B. V., 1993, Influence of the dipolar magnetic field on transport of proton-H atom fluxes in the atmosphere, *Ann. Geophys.*, *11*, 697.

Kunches, J. M, and R. D. Zwickl, 1999, The effects of coronal holes on the propagation of solar energetic protons, *Radiat. Meas.*, *30*, 281.

Lanchester, B. S., J. R. Palmer, M. H. Rees, D. Lummerzheim, K. Kaila, and T. Turunen, 1994, Energy flux and characteristic energy of an elemental auroral structure, *Geophys. Res. Lett.*, *21*, 2789.

Lanchester, B. S., M. H. Rees, D. Lummerzheim, A. Otto, H. U. Frey, and K. U. Kaila, 1997, Large fluxes of auroral electrons in filaments of 100m width, *J. Geophys. Res.*, *102*, 9741.

Lanchester, B. S., M. Galand, S. C. Robertson, M. H. Rees, D. Lummerzheim, I. Furniss, L. M. Peticolas, H. U. Frey , J. Baumgardner, and M. Mendillo, 2003, High resolution measurements and modeling of auroral hydrogen emission line, *Ann. Geophys.*, *21*, 1.

Link, R., 1992, Feautrier solution of the electron transport equation, *J. Geophys. Res.*, *97*, 159.

Liou, K., P. T. Newell, C.-I. Meng, M. Brittnacher, and G. Parks, 1997, Synoptic auroral distribution: A survey using Polar ultraviolet imagery, *J. Geophys. Res., 102*, 21, 197.

Lorentzen, D. A., F. Sigernes, and C. S. Deehr, 1998, Modeling and observations of dayside auroral hydrogen emission Doppler profiles, *J. Geophys. Res.*, *103*, 17, 479.

Lummerzheim, D., and J. Lilensten, 1994, Electron transport and energy degradation in the ionosphere: Evaluation of the numerical solution, comparison with laboratory experiments and auroral observations, *Ann. Geophys.*, *12*, 1039.

Lummerzheim, D., M. Brittnacher, D. Evans, G. A. Germany, G. K. Parks, M. H. Rees, and J. F. Spann, 1997, High time resolution study of the hemispheric power carried by energetic electrons into the ionosphere during the May 19/20, 1996 auroral activity, *Geophys. Res. Lett.*, *24*, 987.

Lummerzheim, D., and M. Galand, 2001, The profile of the hydrogen Hβ emission line in proton aurora, *J. Geophys. Res.*, *106*, 23.

Lummerzheim, D., M. Galand, J. Semeter, M. J. Mendillo, M. H. Rees, F. J. Rich, 2001, Emission of OI(630 nm) in proton aurora, *J. Geophys. Res.*, 106, 141.

McEwen, D. J., and D. A. Harrington, 1991, Polar airglow and aurora, *Can. J. Phys., 69,* 1055.

Mende, S.B., H. Heetderks, H.U. Frey, M. Lampton, S.P. Geller, S. Habraken, E. Renotte, C. Jamar, P. Rochus, J. Spann, S.A. Fuselier, J.-C. Gerard, R. Gladstone, S. Murphree, L. Cogger, 2000, Far Ultraviolet Imaging from the IMAGE Spacecraft: 1. System Design, Space Sci. Rev., IMAGE Special Issue, Vol. *91*, 243.

Min, Q. L., D. Lummerzheim, M. H. Rees, and K. Stamnes, 1993, The Effects of a Parallel Electric Field and the Geomagnetic Field in the Topside Ionosphere on Auroral and Photoelectron Energy Distributions, *J. Geophys. Res.*, *98*, 19,223.

Nagy, A. F., and P. M. Banks, 1970, Photoelectron fluxes in the ionosphere, *J. Geophys. Res.*, *75*, 6260.

Newell, P. T. and C.-I. Meng, 1988, The cusp and the cleft/boundary layer: Low-altitude identifications and statistical local time variation, *J. Geophys. Res.*, *93*, 14549-14556.

Newell, P. T. and C.-I. Meng, 1990, Intense keV energy polar rain, *J. Geophys. Res.*, 95, 7869.

Newell, P. T., S. Wing, C.-I. Meng, and V. Sigillito, 1991, The auroral oval position, structure, and intensity of precipitation from 1984 onwards: An automated online data base, *J. Geophys. Res.*, *96,* 5877-5882.

Newell, P. T., C.-I. Meng, and K. M. Lyons, 1996, Suppression of discrete aurora by sunlight, *Nature,* 381, 766.

Peticolas, L. M., and D. Lummerzheim, 2000, Time-dependent Transport of Field-aligned Bursts of Electrons in Flickering Aurora, *J. Geophys. Res.*, *105*, 12, 895.

Picone, J. M., A. E. Hedin, D. P. Drob, and A. C. Aikin, 2002, NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues, *J. Geophys. Res.*, DOI 10.1029/2002JA009430.

Rees, M. H., D. Lummerzheim, R. G. Roble, J. D. Winningham, J. D. Craven, and L. A. Frank, 1988, Auroral Energy Deposition Rate, Characteristic Electron Energy, and Ionospheric Parameters Derived from Dynamics Explorer 1 Images, *J. Geophys. Res.*, *93*, 12841.

Rodger, C. J., M. A. Clilverd, P. T. Verronen, T. Ulrich, M. J. Jarvis, and E. Turunen, 2006, Dynamic geomagnetic rigidity cutoff variations during a solar proton event, *J. Geophys. Res.*, DOI 10.1029/ 2005JA011395.

Robinson, R. M. and R. R. Vondrak, 1985, Characteristics and sources of ionization in the continuous aurora, *Radio Sci., 20*, 447.

Solomon, S. C., 1993, Auroral electron transport using the Monte Carlo method, *Geophys. Res. Lett.*, *20*, 185.

Stamnes, K., 1980, Analytic approach to auroral electron transport and energy degradation, *Planet. Space Sci.*, *28*, 427.

Stamnes, K., 1981, On the two-stream approach to electron transport and thermalization, *J. Geophys. Res.*, *86*, 2405.

Stenbaek-Nielsen, H. C., T. J. Hallinan, D. L. Osborne, J. Kimball, C. Chaston, J. McFadden, G. Delory, M. Temerin, and C. W. Carlson, 1998, Aircraft observations conjugate to FAST of auroral arc thicknesses, *Geophys. Res. Lett., 25*, 2073.

Stephan, A.W., S. Chakrabarti, and D. M. Cotton, 2000, Evidence of ENA precipitation in the EUV dayglow, *Geophys. Res. Lett., 27*, 2865.

Stephan, A.W., K.F. Dymond, S.A. Buszien, S. E. Thonnard, and R.P. McCoy, 2004, Oxygen aurora during the recovery phase of a major geomagnetic storm, *J. Geophys. Res., 109*, A09208.

Strickland, D. J., D. L. Book, T. P. Coffey, and J. A. Fedder, 1976, Transport equation techniques for the deposition of auroral electrons, *J. Geophys. Res.*, *81*, 2755.

Strickland, D.J., R. Link, and L.J. Paxton, 1992, Far UV remote sensing of thermospheric composition, *Ultraviolet Technology IV, SPIE vol. 1764,* 117-131.

Strickland, D. J., R. E. Daniell Jr., J. R. Jasperse, and B. Basu, 1993, Transport-theoretic model for the electron-proton-hydrogen atom aurora, 2, model results, 1993, *J. Geophys. Res.*, *98*, 21,533.

Synnes, S. A., F. Søraas, and J. P. Hansen, 1998, Monte-Carlo simulation of proton aurora, *J. Atmos. Sol.-Terr. Phys.*, *60*, 1695.

Tanaka, Y.-M., M. Ishii, M. Kubota, Y. Monzen, Y. Murayama, H. Mori, and D. Lummerzheim, 2005, Two-dimensional distribution of medium-energy precipitating electrons estimated from optical and CNA measurements, *Ann. Geophys., 23*, 1555.

Winningham, J. D., and W. J. Heikkila, 1974, Polar cap auroral electron fluxes observed with ISIS 1, *J. Geophys. Res.*, *79*, 949.

Winningham, J. D., J. R. Sharber, R. A. Frahm, J. L. Burch, N. Eaker, R. K. Black, V. A. Belvins, J. P. Andrews, J. Rudzki, D. L. Chenette, D. W. Datlowe, E. E. Gaines, W. I. Imhof, R. W. Nightingale, S. S. Prasad, R. M. Robinson, T. L. Schumaker, R. R. Vondrak, P. F. Bythrow, B. J. Anderson, T. A. Potemra, L. J. Zanetti, D. B. Holland, M. H. Rees, D. Lummerzheim, G. C. Reid, R. G. Roble, C. R. Clauer, and P. M. Banks, 1993, The UARS Particle Environment Monitor, *J. Geophys. Res.*, *98*, 10, 649.

1.3.3   Ionospheric Plasma Convection

(Mike Ruohoniemi, The John Hopkins University Applied Physics Laboratory)

1.3.3.1   Introduction

The ionosphere is the lowest level of the atmosphere that displays significant electrical properties. Even so, the ionized component constitutes only a very small percentage of the total mass. In the absence of electrical and magnetic fields the motion of the charged particles would be dictated by the dynamics of the neutral atmosphere. With the imposition of a magnetic field alone, the motion of the charged particles becomes subject to the q$\mathbf{v}\times\mathbf{B}$, or Lorentz, force, which compels them to gyrate about field lines. When an electric field is added, the particles acquire a drift in the $\mathbf{E}\times\mathbf{B}$ direction irrespective of particle charge. The motion of ionospheric plasma under the influence of electric field perpendicular to $\mathbf{B}$ constitutes convection. At high latitudes the geomagnetic field lines are almost vertical and so the plasma convection is effectively confined to the horizontal plane. At their upper ends the field lines are in contact with large-scale electric fields that are generated by the flow of the solar wind past the Earth. These fields map with

little attenuation along the magnetic field lines into the ionosphere. The $\mathbf{E}\times\mathbf{B}$ motion of plasma in the outer magnetosphere is thus projected onto a high-latitude ionospheric 'screen'. The same connection provides a path for energetic particles to precipitate from the magnetosphere and cause the aurora. Both the large-scale ionospheric precipitation and convection can be viewed advantageously within a reference frame that is centered on the geomagnetic pole and extends down in latitude to an auroral boundary that is typically at 60 deg Λ.

Figure 1-8 shows a model of convection as envisioned originally by *Dungey* [1961]. The solid lines represent contours of constant electrostatic potential, Φ, to which the electric field is related via $\mathbf{E} = -\text{grad}\Phi$. The plasma circulates along the contours in a manner analogous to the movement of air along isobars of constant pressure in conventional weather maps. The large-scale pattern of convection consists of two cells with antisunward flow over the polar cap closed by return flows along the dawn and dusk flanks. *Dungey* [1961] hypothesized that such a circulation would result from a process on the dayside that merges solar and geomagnetic magnetic field lines with an accompanying transfer of energy and momentum into the magnetosphere. *Axford and Hines* [1961] suggested that a viscous interaction takes place between the solar wind and magnetospheric plasma along the magnetospheric flanks; a similar pattern of large-scale circulation is obtained.
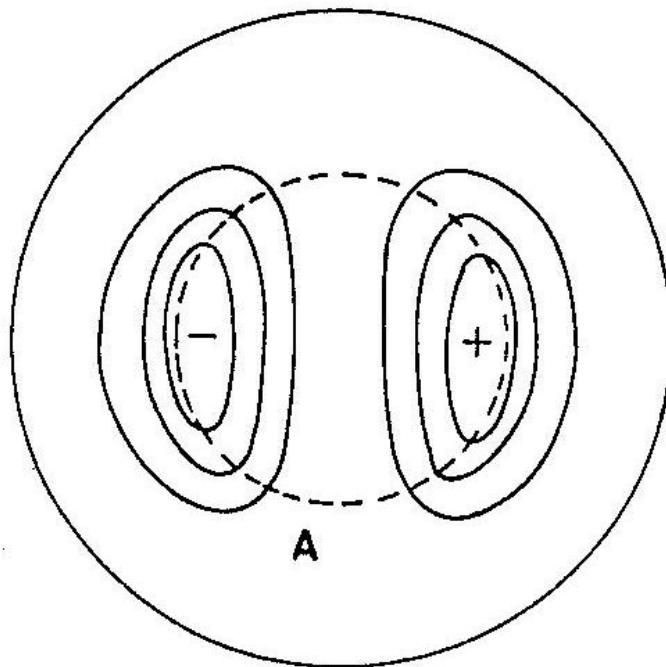
Figure 1-8: Theoretical Pattern of Ionospheric Plasma Convection as Proposed by *Dungey* [1961].

At altitudes within the *E* region (90-150 km), the electrons partake fully in $\mathbf{E}\times\mathbf{B}$ convection but the ions are partially or fully immobilized by collision with neutrals. The velocity difference gives rise to an electrical current. The ionosphere functions as a resistive load on the magnetospheric generators of convection because collisions between charged particles and neutrals dissipate energy through heating. Convection also redistributes ionization and plasma structure throughout the convection zone. The velocities associated with high-latitude convection are typically hundreds of meters per second, often exceed 1500 m/s, and can reach 3-4 km/s. The redistribution of plasma under the control of convection electric field takes place on time scales of hours to days.

1.3.3.2   Methods of Measuring Ionospheric Plasma Convection

A variety of methods have been applied to study the $\mathbf{E}\times\mathbf{B}$ drift of ionospheric plasma. Some of the first direct observations were made with rockets, e.g., *Haerendel et al.* [1967]. These released clouds of barium gas that rapidly ionized and then drifted at the convection velocity. In situ measurements are now more commonly obtained from satellites that carry instruments to detect ion drift. These have included Ogo 6 [*Heppner*, 1972], Atmosphere Explorer C [*Heelis*, 1984], and Dynamics Explorer 2 [*Heelis et al.*, 1986]. The DMSP series of low-altitude satellites [*Hairston and Heelis*, 1995] in particular have provided nearly continuous observations for many years. Ionospheric convection can also be monitored from the ground using remote sensing techniques. Powerful incoherent-scatter radars generate backscatter from thermal fluctuations in the ambient plasma and analysis of the spectra yields information on the $\mathbf{E}\times\mathbf{B}$ drift as well as on many other ionospheric parameters [*Banks et al.*, 1973]. A less direct ground-based method utilizes magnetometers. These instruments detect magnetic disturbances at the ground that can be related to the currents flowing in the *E* region ionosphere. The convection is obtained by deconstructing the current signatures into electric field and conductivity components [*Richmond et al.*, 1988].

Another radar method involves coherent backscatter from irregularities in the ionospheric ionization that have been amplified far beyond thermal levels by plasma instability processes. Much lower levels of power are required than is the case for incoherent scatter but the technique does require that irregularities exist and that the radar signal encounters them while propagating nearly orthogonal to the magnetic field lines. The Doppler shift in the backscattered signal characterizes one component of the $\mathbf{E}\times\mathbf{B}$ drift. One early system operated on VHF frequencies in Scandinavia and used pairs of receivers to resolve stereoscopically the two-dimensional convection velocity vector [Greenwald et al., 1978]. The geometry of the geomagnetic field restricts VHF operations to the E region ionosphere. A system based on operations at HF frequencies was subsequently developed [Greenwald et al., 1985]. The advantage of HF is that the vertical density gradients in the ionosphere can refract the transmissions to propagate orthogonally to geomagnetic field lines at F region altitudes (150-1000 km); this vastly increases the area that is accessible to coherent scatter. An international collaboration known as SuperDARN was assembled to pursue development of chains of HF radars in both the northern and southern hemispheres [Greenwald et al., 1995a].

Figure 1-9 shows the sites and fields of view of the radars as of early 2005. For the most part the radars have been organized into pairs with overlapping fields of view to facilitate stereoscopic observation of the convection velocity. Data from the radar chains are routinely merged with statistical model information and solved for a best-fit image of the large-scale convection pattern in the high-latitude ionosphere [Ruohoniemi and Baker, 1998; Shepherd and Ruohoniemi, 2000]. A 'nowcast', or near real-time specification, of the convection pattern in the northern hemisphere is generated using internet links to the northern radars [Ruohoniemi et al., 2001] and posted to a public access web site (http://superdarn.jhuapl.edu/). The images are generated continuously at a cadence of 1 or 2 minutes. Figure 1-10 shows an example of a convection map that was derived for January 21, 2001, 1540-1542 UT. Velocity vectors are drawn where direct measurements were available from at least one of the radars. As indicated, the prevailing IMF was predominantly southward and a two-cell pattern was obtained. The total variation in the electrostatic potential across the polar cap, ΦPC, was 67 kV. The value of ΦPC typically ranges from 15 to 300 kV, with the lower end encountered for prolonged periods of northward IMF that generally lead to geomagnetic quiescence while the higher end is associated with very large geomagnetic disturbances.
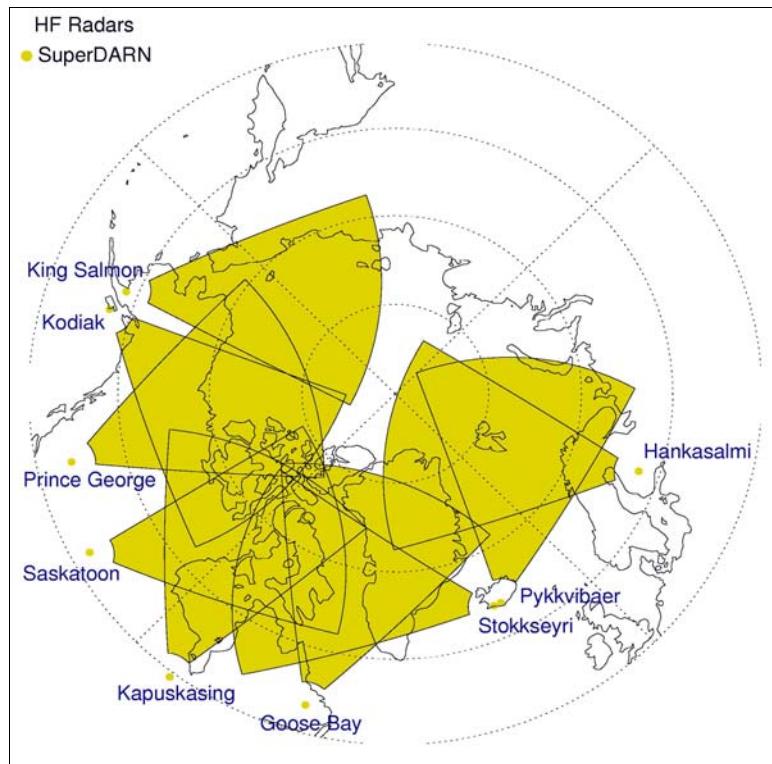
Figure 1-9: Fields of View of the Northern Component of the Super Dual Auroral Radar Network as of Early 2005.
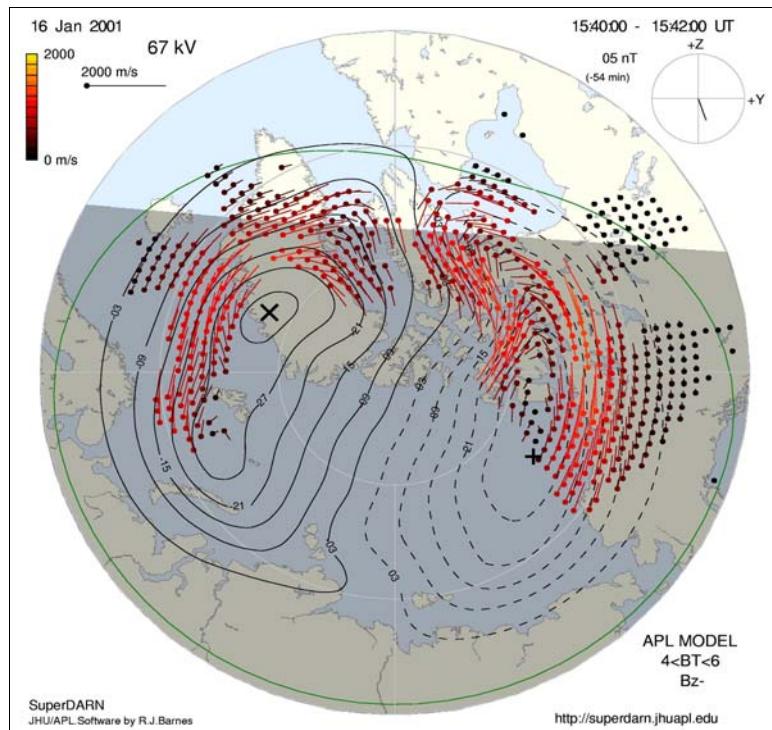
Figure 1-10: An Example of a Map of Ionospheric Plasma Convection Derived from Observations with the SuperDARN HF Radars as Described in the Text. The circular inset map in the upper right hand corner indicates the orientation of the IMF in the GSM *y-z* plane.

1.3.3.3 Dynamics of Plasma Convection

The convection of ionospheric plasma at high latitudes most often conforms to a two-cell pattern of circulation with antisunward flow over the polar cap. It is thought that the merging process envisioned by *Dungey* [1961] usually accounts for more of the convection than the viscous interaction mechanism advanced by *Axford and Hines* [1961]. The essential factors in setting the convection morphology is that geomagnetic field lines merge on the dayside magnetopause and are then subject to an antisunward tug due to their projections into the solar wind. As shown in Figure 1-11, the process of merging, also called reconnection, transforms a geomagnetic field line that was 'closed' between the hemispheres into a field line that 'opens' into interplanetary space. The field line is dragged across the pole and the plasma attached to the line shares in this motion. Open magnetic flux accumulates in the polar cap or is returned to a closed condition by nightside reconnection processes that can include the explosive auroral substorm. The cycle is completed by flows along the dawn and dusk flanks that return closed flux and plasma to the dayside.
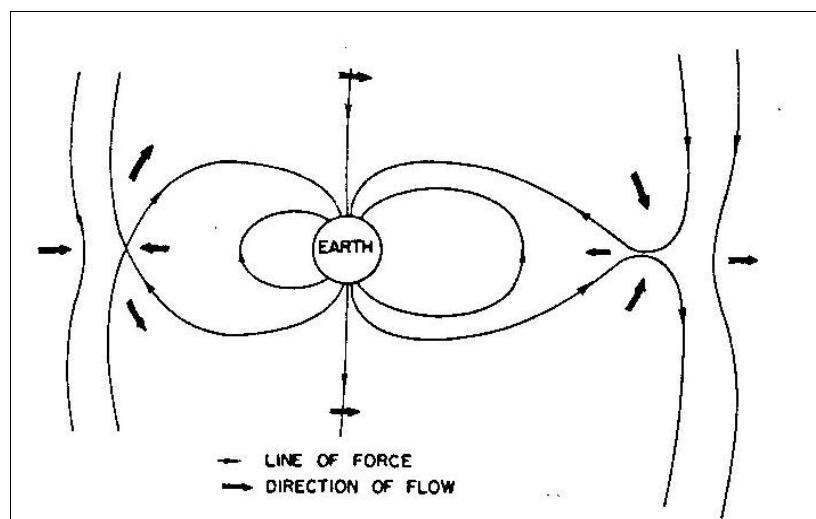
Figure 1-11: Schematic of the Process whereby Geomagnetic Lines on the Dayside Merge with Southward Interplanetary Magnetic Field. The open field lines are carried tailward by the solar wind plasma and are eventually returned to closed dipolar form by merging across the magnetotail – from *Dungey* [1961].

A vital aspect of the merging process is the orientation of the solar wind field line at the magnetopause; merging is favored when the IMF and geomagnetic field lines are anti-parallel. As the geomagnetic field line connects between the hemispheres in a south-to-north sense, *Dungey* [1961] proposed his merging scenario for a southward orientation of the IMF. Later, he considered the possibility of merging for northward IMF [*Dungey*, 1963]. The geomagnetic field at the magnetopause converges at a point and forms a funnel-like feature called the cusp that descends into the high-latitude ionosphere. The convoluted geometry of the cusp field lines combined with tilting of the Earth leads to anti-parallel merging sites even for northward IMF. However, in this case the initial tug felt by the merged field lines is in the sunward direction owing to the stresses on the highly kinked field lines. Eventually the motion of the solar wind dominates and the flow turns antisunward. In the vicinity of the cusp, two small convection cells form that cycle plasma in the reverse sense, i.e., sunward at noon and antisunward along the dawn and dusk flanks, with sunward return flow from midnight to noon. When coupled with weak viscous cell convection of the usual sense, a four-cell pattern of circulation is obtained. If a cell is very weak it may practically disappear, leaving a three-cell pattern.

Other merging scenarios for driving high-latitude convection are possible (see *Crooker* [1992] for a review). For example, within the polar cap, the open field lines may themselves present merging sites to

the incoming IMF, leading to open-to-open conversions where a geomagnetic field line exchanges its extension into the solar wind in its own hemisphere for a line that overdrapes the magnetopause on the dayside and extends into the solar wind in the opposite hemisphere. The resulting plasma motion, which is reversed from the usual two-cell sense, amounts to a continual stirring of open field lines within the polar cap and is referred to as lobe cell convection. The character of the global convection (two-cell, distorted two-cell, multi-cell, reverse, etc.) is observed to hinge on the relative magnitudes of the north-south and azimuthal components of the IMF, e.g., *Knipp et al.* [1993].

These considerations indicate the paramount importance that attaches to the solar wind, especially the IMF, in determining plasma convection. Many measurement techniques have been applied to study the nature of the IMF effect. These have lead to numerous statistical studies that describe the average pattern of ionospheric plasma convection as a function of IMF, e.g., *Heppner and Maynard* [1987], *Holt et al.* [1987], *Papitashvili et al.* [1994], *Weimer* [1995], *Ruohoniemi and Greenwald* [1996]. Figure 1-12 shows an example from *Ruohoniemi and Greenwald* [1996] that was derived on the basis of measurements from the prototype SuperDARN HF radar located at Goose Bay, Labrador. Note the intensification of flows and the enlargement of the two cell pattern for increasingly southward IMF and the emergence of more convoluted dayside flows for increasingly northward IMF. The azimuthal (east-west) component of the IMF is seen to play a major role in orienting the dayside flows and defining the shapes of the cells. Figure 1-13 shows the average pattern derived for strongly northward IMF conditions; two small but distinct reverse dayside convection cells emerge bounded at lower latitude by cells with the usual sense of circulation. It turns out that the various studies are remarkably consistent in their depictions of the average IMF dependencies. The statistical characterization of the convection pattern in terms of the IMF can be considered solved and interest has moved on to the impact of secondary factors, such as season, e.g., *Ruohoniemi and Greenwald* [2005].
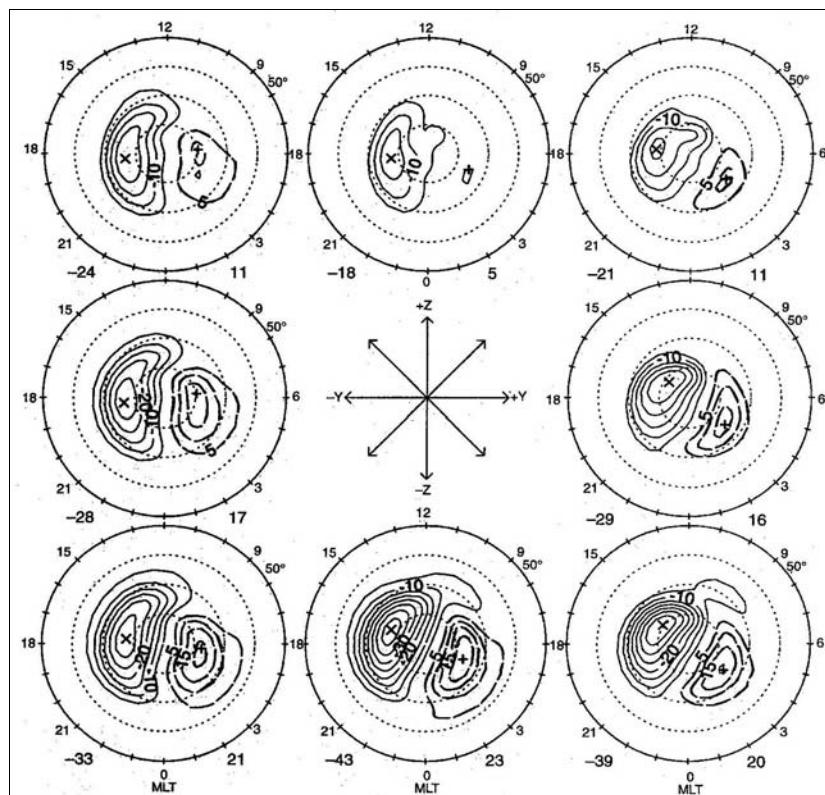
Figure 1-12: Average Patterns of Ionospheric Plasma Convection Sorted by the Orientation of the IMF in the GSM y-z Plane for a Moderate Level of Geomagnetic Disturbance. The values of the potential extrema are indicated (in kV) – from *Ruohoniemi and Greenwald* [1996].
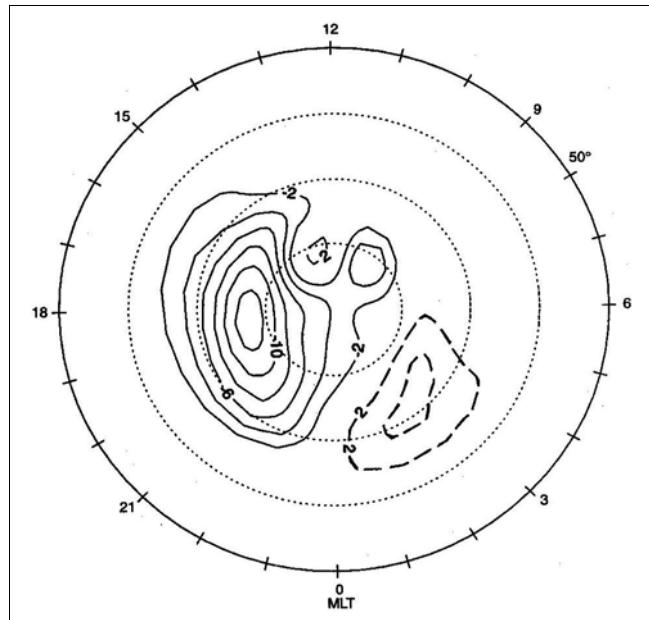
Figure 1-13: Average Pattern of Ionospheric Plasma Convection Derived for Strongly Northward IMF − from *Ruohoniemi and Greenwald* [1996].

It has been more challenging to study and understand the instantaneous convection pattern, i.e., that which prevails in the ionosphere at an instant in time. It has not been possible to make observations everywhere all at once although the incorporation of individual instruments into larger networks such as SuperDARN is greatly increasing our capabilities. The first large-scale depictions of instantaneous convection patterns were generated with the AMIE technique which relies on the inversion of magnetometer data [*Richmond and Kamide*, 1988]. These illustrated that the instantaneous convection was quite variable and sensitive to IMF variations, e.g., *Knipp et al.* [1991]. An early triumph of the SuperDARN concept was the first direct mapping of a reverse dayside convection cell for northward IMF [*Greenwald et al.*, 1995*b*]. With the continuing improvement in coverage, the nature of the convection response to changed IMF came under intense scrutiny. It had been argued that the initial response in ionospheric convection would be confined to the dayside cusp region where the plasma has the most direct connection to conditions at the magnetopause and then progress towards the nightside with delays measured in the tens of minutes to hours, reflecting the relatively slow progress of convecting flux tubes. However, observations with distributed sets of magnetometers and HF radars indicated that the convection begins to respond almost simultaneously (within minutes) at all longitudes [*Ridley et al.*, 1998; *Ruohoniemi et al.* 1998]. A lively controversy ensued [*Lockwood*, 1999; *Ridley et al.*, 1999]. Continued observations with the expanding instrument networks decided the issue in favor of simultaneous onset [*Murr and Hughes*, 2001; *Ruohoniemi et al.*, 2002]. The manner in which ionospheric convection evolves following the onset of a response remains an active study area, e.g*., Lu et al.* [2002].

Plasma convection is in general a highly variable phenomenon. Part of the variability is attributable to variability in the solar wind driver. The magnitude and direction of the IMF vary typically on time scales of hours but can occur on steps as short as one minute. Advance knowledge of the IMF gives a basis for predicting the convection using one of the IMF-keyed statistical models discussed earlier. Unfortunately, the most advance warning that can be managed currently is about 1 hour. This is set by the propagation time of the solar wind from the position of the ACE satellite at the Lagrangian L1 point to effects at the magnetopause. Furthermore - and very significantly for current directions in research - the convection is not completely determined by the IMF or solar wind factors. Marked variability in convection for stable IMF has been noted in both fine-scale [*Codrescu et al.*, 2000; *Shepherd et al.*, 2003] and large-scale

[*Matsuo et al.*, 2003; *Bristow et al.*, 2004] observations. Indeed, the ability of the standard statistical convection models to predict the instantaneous local convection is not very high, e.g., *Eriksson et al.*, [2002]. This likely points to the influence of factors internal to the magnetosphere-ionosphere (M-I) system. It is perhaps not surprising that, as an auroral phenomenon, ionospheric plasma convection at high latitudes exhibits pronounced variability in time and space. There are notable impacts on convection during auroral substorm activity, e.g., Provan et al. [2004] that appear to be a consequence of balancing electric fields and conductivities in the ionosphere with currents supplied by the magnetosphere. Understanding the origins and effects of the variability in convection is a very active research area and presents a particular challenge for efforts to model the coupling of the solar wind to the M-I system using MHD codes.

1.3.3.4   Conjugacy

The northern and southern hemispheres are linked by the lines of force of the geomagnetic field. This coupling is significant as charged particles move relatively freely along the field lines and so auroral processes tend to mirror between the hemispheres in regions that are magnetically conjugate. Ionospheric convection also exhibits conjugacy in that the electrostatic potential, Φ, associated with the large-scale convection electric fields maps between the hemispheres along the geomagnetic field lines. Thus convection in the northern hemisphere should mirror in the southern hemisphere within the region of closed field lines. At the higher latitudes associated with the polar cap the field lines are open to the IMF in the solar wind and interhemispheric conjugacy cannot be expected.

The mapping of geomagnetic flux between the hemispheres varies with season and UT and reflects the secular variation in the geomagnetic field, e.g., *Stassinopoulos et al.*, [1984]. Surprisingly perhaps, given the long history of observations in Antarctica, the extent of conjugate behaviour in auroral phenomenon has still not been fully worked out. This is especially true of ionospheric plasma convection. Of course, proper conjugacy is limited to field lines that close between the hemispheres and convection dynamics are largely controlled by processes that involve the transformation of field lines between closed and open. There is also the difficulty of determining which ionospheric plasma parcels are truly conjugate as the geomagnetic field lines can depart dramatically from their nominal dipolar geometry especially in the vicinity of the open/closed field line boundary.

An overall view of the relation between convection in the northern and southern hemispheres can be obtained by comparing convection patterns. It was noted by *Heppner and Maynard* [1987] that the dayside ionospheric flows conform to signatures ordered by the sign of the azimuthal component of the IMF, i.e., B_y, and that the convection in the northern hemisphere for B_y+ resembles the convection in the southern hemisphere for B_y- and vice versa. The first statistical convection model derived exclusively for the southern hemisphere was based on observations with an HF radar located at Halley Station, Antarctica [*Leonard et al.*, 1995]. Comparison with the northern observations confirms that the large-scale convection pattern is asymmetric between the hemispheres with respect to the sign of B_y. This is related to the anti-parallel condition for merging between the IMF and the geomagnetic field at the magnetopause and the reversal in the direction of the azimuthal component of the geomagnetic field lines between the hemispheres. The B_y-sign factor was demonstrated in instantaneous convection by *Greenwald et al.* [1990], who observed the emergence of distinctive cusp flows in the two hemispheres following changes in the sign of B_y using observations from the Goose Bay and Halley Station HF radars. Figure 1-14 shows an example. Flows on the dayside were oriented more westward (dawnward) in the northern/southern hemisphere for B_y+/B_y- conditions but acquired a pronounced shear in latitude when the sign of B_y subsequently reversed. An analogous interhemispheric asymmetry in nightside convection was recently reported by *Grocott et al.* [2005].
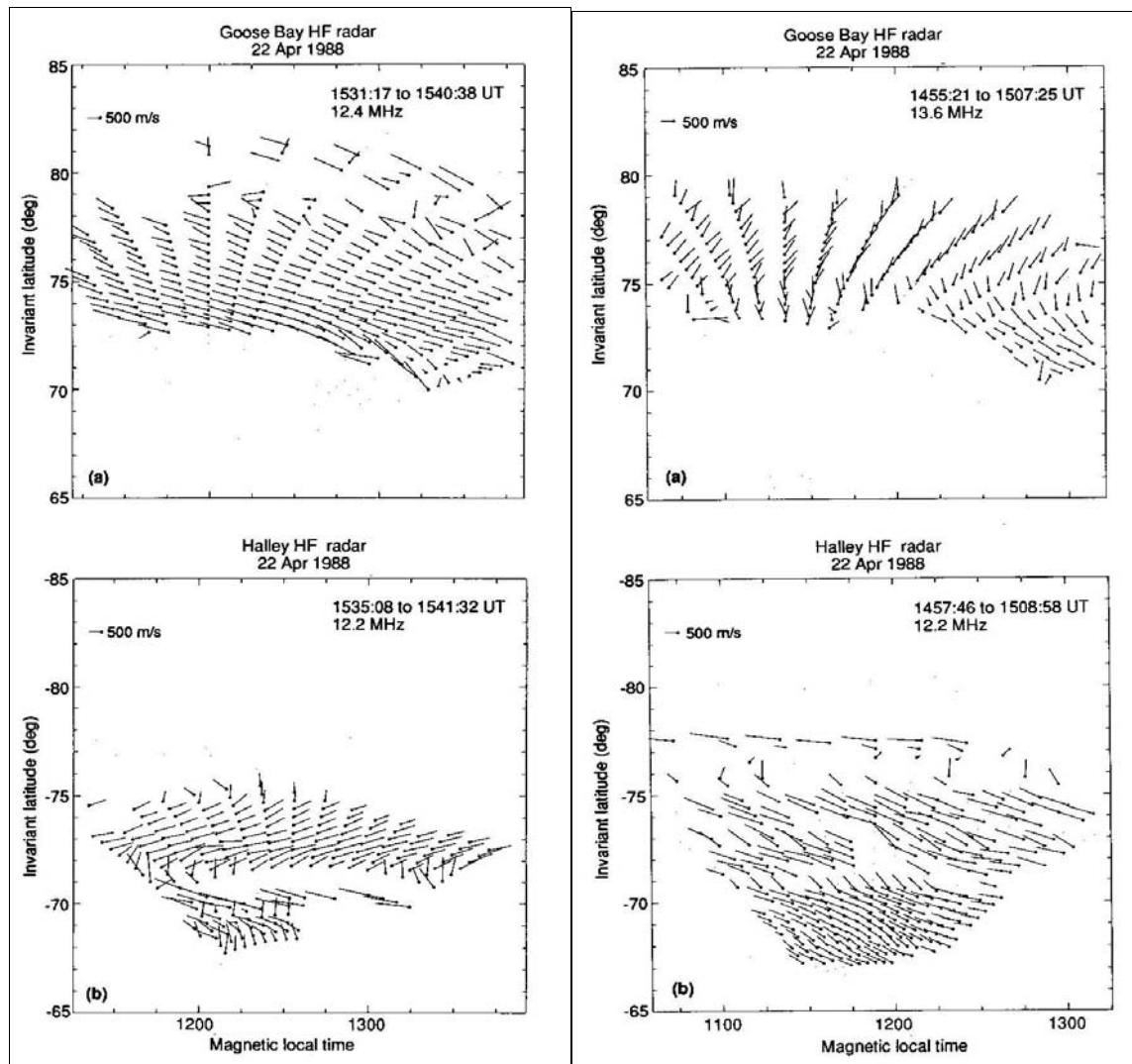
**Figure 1-14: Instantaneous Patterns of Dayside Plasma Convection Obtained
in Conjugate Hemispheres for Distinct B_y Conditions.**

For B_y+, the northern radar (Goose Bay) observes westward flow while the southern radar (Halley) observes a shear reversal in latitude. For subsequent B_y- conditions, the flows in the northern hemisphere are sheared in latitude while those in the southern are westward and poleward – after Greenwald et al. [1991].

1.3.3.5   Impacts of Convection

Convection has a range of significant effects in the high-latitude ionosphere. At F region altitudes (> 200 km) the lifetime of ionization is fairly long (~hours) and plasma can be transported large distances with little change. Consequently, the distribution of high-latitude plasma density at any instant represents the cumulative effect of convection acting on the sources of ionization over many hours. Since convection pattern varies on time scales of minutes to hours, redistribution of plasma density can be expected over a range of spatial scales and there are many possibilities for mixing dissimilar plasma populations to create structure. Figure 1-15 shows a particularly dramatic example of structuring due to *Foster et al.* [2005] who showed in combined GPS TEC/convection observations the propagation of a storm-enhanced density feature (SED) from mid to high latitudes during a major geomagnetic storm. Because of convection the ionization overhead at a location deep on the nightside can in fact be due to processes that took place hours earlier on the dayside.
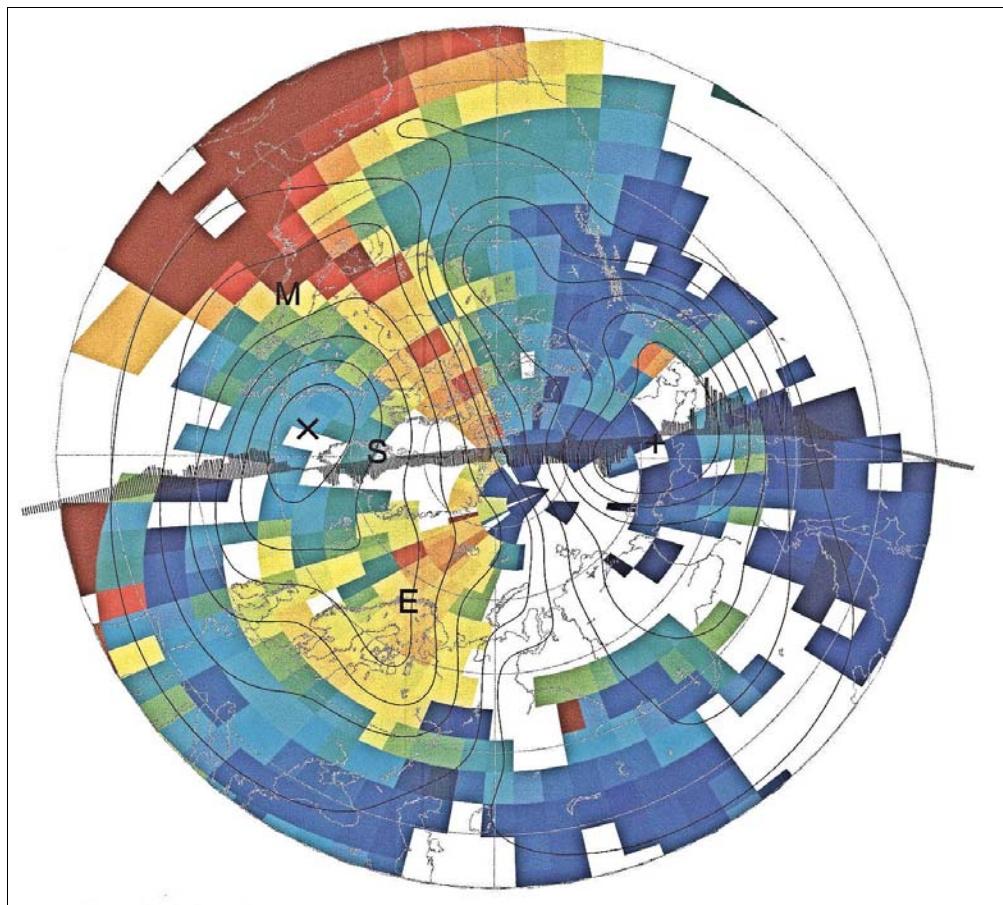
Figure 1-15: Map of GPS Total Electron Content (TEC) Data in Polar Projection Showing a Storm-Enhanced Density Plume Extending Continuously from a Low-Latitude Source Region in the Postnoon Sector through the Polar Cap into the Midnight Sector.

The locations of large incoherent scatter radar facilities that observed related effects are shown by the letters (M: Millstone Hill, S: Sondrestrom, E: EISCAT). Overplotted are convection contours deduced from combined SuperDARN radar and DMSP satellite observations – after *Foster et al.* [2005].

Large-scale models of the ionosphere attempt to account comprehensively for the condition of the high-latitude ionosphere by incorporating the sources of ionization, convection, and chemical effects in numerical constructs with long timelines, e.g., the Utah State University (USU) model described by *Schunk and Sojka* [1987]. The evolution of enhancements in ionospheric ionization known as patches has been modeled. These features form on the dayside by processes that may involve convection dynamics (e.g., *Valladares et al.* [1994]) and are transported by convection across the polar cap to the nightside where they become elongated in longitude as determined by the convection trajectories and form what are known as boundary blobs, e.g., *Anderson et al.*, [1996], *Crowley et al.* [2000]. There are currently major efforts to model the global ionosphere in near real-time using, among other inputs, as much information on convection as can be obtained, e.g., *Schunk et al.* [2004].

In addition to transporting bulk ionospheric plasma, the electric fields associated with convection contribute to plasma instabilities that generate irregularities in the ionization. Most often the irregularities are thought of as waves in the plasma that range in scale from tens of kilometers to centimetres. Other factors that contribute to the formation of irregularities include gradients in plasma density and particle temperatures (see reviews by *Fejer and Kelley* [1980] and *Keskinen* [1983]). The irregularities can be observed with

radars that operate on wavelengths that generate Bragg scattering from the plasma waves, generally on frequencies from HF to UHF. For conventional surveillance and communications radars the irregularities cause clutter that can severely impair operations. GPS systems experience phase fluctuations when transiting affected volumes and their performance is degraded. Efforts to understand the interplay of convection and plasma structure in generating irregularities are becoming more sophisticated with the increasing availability of high powered computer systems for numerical modeling, e.g., Guzdar et al. [2001].

1.3.3.6   References

Anderson, D.N., D.T. Decker, and C.E. Valladares, Modeling boundary blobs using time varying convection, *Geophys. Res. Lett.*, *23*, 579-582, 1996.

Axford, W.I., and C.O. Hines, A unifying theory of high-latitude geophysical phenomena, *Canadian Journal of Physics*, *39*, 1433-1464, 1961.

Baker, J.B.H., R.A. Greenwald, J.M. Ruohoniemi, M. Forster, G. Paschmann, E.F. Donovan, N.A. Tsyganenko, J.M. Quinn, and A. Balogh, Conjugate comparison of Super Dual Auroral Radar Network and Cluster electron drift instrument measurements of ExB plasma drift, *Journal of Geophysical Research*, *109*, doi: 10.1029/2003JA009912, 2004.

Banks, P.M., J.R. Doupnik, and S.I. Akasofu, Electric field observations by incoherent scatter in the auroral zone, *J. Geophys. Res.*, *78*, 6607, 1973.

Bristow, W.A., R.A. Greenwald, S.G. Shepherd, and J.M. Hughes, On the observed variability of the cross-polar cap potential, *J. Geophys. Res.*, *109*, A02203, doi: 10.1029/2003JA010206, 2004.

Codrescu, M.V., T.J. Fuller-Rowell, J.C. Foster, J.M. Holt, and S.J. Cariglia, Electric field variability associated with the Millstone Hill electric field model, *Journal of Geophysical Research*, *105*, 5265-5273, 2000.

Crowley, G., A.J. Ridley, D. Deist, S. Wing, D.J. Knipp, B.A. Emery, J. Foster, R. Heelis, M. Hairston, and B.W. Reinisch, Transformation of high-latitude ionospheric F region patches into blobs during the March 21, 1990, storm, *J. Geophys.l Res.*, *105*, 5125-5230, 2000.

Dungey, J.W., Interplanetary magnetic fields and the auroral zones, *Physical Review Letters*, *6*, 47-48, 1961.

Dungey, J. W., The structure of the exosphere or adventures in velocity space, in *Geophysics, The Earth's Environment*, edited by E.C. Dewitt, J. Hieblot, and A. Lebeau, pp. 505-550, Gordon and Breach, New York, 1963.

Eriksson, S., L.G. Blomberg, and D.R. Weimer, Comparing a spherical harmonic model of the global electric field distribution with Astrid-2 observations, *Journal of Geophysical Research*, *107*, 1391, doi:10.1029/2002JA009313, 2002.

Fejer, B.G., and M.C. Kelley, Ionospheric irregularities, *Rev. Geophysics Space Res.*, *18*, 401, 1980.

Foster, J.C., A. J. Coster, P. J. Erickson, J. H. Holt, F. D. Lind, W. Rideout, M. McCready, A. van Eyken, R. J. Barnes, R. A. Greenwald, and F. J. Rich, Multiradar observations of the polar tongue of ionization, *Journal of Geophysical Research*, *110*, doi: 10.1029/2004JA010928, 2005.

Greenwald, R.A., W. Weiss, E. Nielsen, and N.R. Thomas, STARE: a new radar auroral backscatter experiment in northern Scandinavia, *Radio Science*, *12*, 1021-1039, 1978.

Greenwald, R.A., K.B. Baker, R.A. Hutchins, and C. Hanuise, An HF phased-array radar for studying small-scale structure in the high-latitude ionosphere, *Radio Science*, *20*, 63-79, 1985.

Greenwald, R.A., K.B. Baker, J.M. Ruohoniemi, J.R. Dudeney, M.J. Pinnock, N. Mattin, J.M. Leonard, and R.P. Lepping, Simultaneous conjugate observations of dynamic variations in high latitude dayside convection due to changes in IMF B_y, *Journal of Geophysical Research*, *95*, 805708072, 1990.

Greenwald, R.A., K.B. Baker, J.R. Dudeney, M. Pinnock, T.B. Jones, E.C. Thomas, J.-P. Villain, J.-C. Cerisier, C. Senior, C. Hanuise, R.D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A.D.M. Walker, N. Sato, and H. Yamagishi, DARN/SuperDARN: A global view of high-latitude convection, *Space Science Reviews*, *71*, 763-796, 1995.

Greenwald, R.A., W.A. Bristow, G.J. Sofko, C. Senior, J.-C. Ceriser, and A. Szabo, SuperDual Auroral Radar Network radar imaging of dayside high-latitude convection under northward interplanetary magnetic field: Toward resolving the distorted two-cell versus multicell controversy, *Journal of Geophysical Research*, *100*, 19661-19674, 1995.

Grocott, A., T.K. Yeoman, S.E. Milan, and S.W.H. Cowley, Interhemispheric observations of the ionospheric signature of tail reconnection during IMF-northward non-substorm intervals, *Ann. Geophys.*, *23*, 1763-1770, 2005.

Guzdar, P.N., X. Shao, C.C. Goodrich, K. Papadopoulos, M.J. Wiltberger, and J.G. Lyon, Three-dimensional MHD simulations of the steady state magnetosphere with northward interplanetary magnetic field, *Journal of Geophysical Research*, *106*, 275-287, 2001.

Hairston, M.R., and R.A. Heelis, Response time of the polar ionospheric convection pattern to changes in the north-south direction of the IMF, *Geophysical Research Letters*, *22*, 631-634, 1995.

Haerendel, G.R., R. Lust, and E. Reiser, Motion of artificial ion clouds in the upper atmosphere, *Planetary Space Science*, *15*, 1, 1967.

Heelis, R. A., The effects of interplanetary magnetic field orientation on day side high-latitude convection, *J. Geophys. Res.*, *89*, 2873-2880, 1984.

Heelis, R.A., P.H. Reiff, J.D. Winningham, and W.B. Hanson, Ionospheric convection signatures observed by DE 2 during northward interplanetary magnetic field, *J. Geophys. Res.*, *91*, 5817-5830, 1986.

Heppner, J.P., Polar cap electric field distributions related to the interplanetary magnetic field directions, *J. Geophys. Res.*, *77*, 4877, 1972.

Heppner, J.P., and N.C. Maynard, Empirical high-latitude electric field models, *Journal of Geophysical Research*, *92*, 4467-4489, 1987.

Holt, J.M., R.H. Wand, J.V. Evans, and W.L. Oliver, Empirical models for the plasma convection at high latitudes from Millstone Hill observations, *Journal of Geophysical Research*, *92*, 203-212, 1987.

Keskinen, M.J., and S.L. Ossakow, Theories of high-latitude ionospheric irregularities: A review, *Radio Sci.*, *18*, 1077-1091, 1983.

Knipp, D.J., A.D. Richmond, B. Emery, N.U. Crooker, O.D.L. Beaujardiere, D. Evans, and H. Kroehl, Ionospheric convection response to changing IMF direction, *Geophysical Research Letters*, *18*, 721-724, 1991.

Knipp, D.J., B.A. Emery, A.D. Richmond, N.U. Crooker, M.R. Hairston, J.A. Cumnock, W.F. Denig, F.J. Rich, O.d.l. Beaujardiere, J.M. Ruohoniemi, A.S. Rodger, G. Crowley, B.-H. Ahn, D.S. Evans, T.J. Fuller-Rowell, E. Friis-Christensen, M. Lockwood, H.W. Kroehl, C.G. MacLennan, A. McEwin, R.J. Pellinen, R.J. Morris, G.B. Burns, V. Papitashvili, A. Zaitzev, O. Troshichev, N. Sato, P. Sutcliffe, and L. Tomlinson, Ionospheric convection response to slow, strong variations in a northward interplanetary magnetic field: A case study for January 14, 1988, *Journal of Geophysical Research*, *98*, 19273-19292, 1993.

Leonard, J.M., M. Pinnock, A.S. Rodger, J.R. Dudeney, R.A. Greenwald, and K.B. Baker, Ionospheric plasma convection in the southern hemisphere, *J. Atmos. Terr. Phys.*, 57, 889-897, 1995.

Lockwood, M.A.S.W.H.C., Comment on "A statistical study of the ionospheric convection response to changing interplanetary magnetic field conditions using the assimilative mapping of ionospheric electrodynamics technique" by A. J. Ridley et al., *Journal of Geophysical Research*, *104*, 4387-4391, 1999.

Lu, G., T.E. Holzer, D. Lummerzheim, J.M. Ruohoniemi, P. Stauning, O. Troshichev, P.T. Newell, M. Brittnacher, and G. Parks, Ionospheric response to the interplanetary magnetic field southward turning: Fast onset and slow reconfiguration, *Journal of Geophysical Research*, *107*, 10.1029/2001JA000324, 2002.

Matsuo, T., A.D. Richmond, and K. Hensel, High-latitude ionospheric electric field variability and electric potential derived from DE-2 plasma drift measurements: Dependence on IMF and dipole tilt, *Journal of Geophysical Research*, *108*, 1005, doi:10.1029/2002JA009429, 2003.

Murr, D.L., and W.J. Hughes, Reconfiguration timescales of ionospheric convection, *Geophysical Research Letters*, *28*, 2145-2148, 2001.

Papitashvili, V.O., B.A. Belov, D.S. Faermark, Y.I. Feldstein, S.A. Golyshev, L.I. Gromova, and A.E. Levitin, Electric potential patterns in the northern and southern polar regions parameterized by the interplanetary magnetic field, *Journal of Geophysical Research*, *9*, 13251-13262, 1994.

Provan, G., M. Lester, S.B. Mende, and S.E. Milan, Statistical study of high-latitude plasma flow during magnetospheric substorms, *Annales Geophysicae*, *22*, 3607-3624, 2004.

Richmond, A.D., and Y. Kamide, Mapping electrodynamic features of the high-latitude ionosphere from localized observations: techniques, *J. Geophys. Res.*, *93*, 5741-5759, 1988.

Ridley, A.J., G. Lu, C.R. Clauer, and V.O. Papitashvili, A statistical study of the ionospheric convection response to changing interplanetary magnetic field conditions using the assimilative mapping of ionospheric electrodynamics technique, *Journal of Geophysical Research*, *103*, 4023-4039, 1998.

Ridley, A.J., G. Lu, C.R. Clauer, and V.O. Papitashvili, Reply, *Journal of Geophysical Research*, *104*, 4393-4396, 1999.

Ruohoniemi, J.M., and R.A. Greenwald, Statistical patterns of high-latitude convection obtained from Goose Bay HF radar observations, *Journal of Geophysical Research*, *101*, 21743-21763, 1996.

Ruohoniemi, J.M., and K.B. Baker, Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations, *J. Geophys. Res.*, *103*, 20797-20811, 1998.

Ruohoniemi, J.M., and R.A. Greenwald, The response of high-latitude convection to a sudden southward IMF turning, *Geophysical Research Letters*, *25*, 2913-2916, 1998.

Ruohoniemi, J.M., R.J. Barnes, R.A. Greenwald, and S.G. Shepherd, The response of the high-latitude ionosphere to the coronal mass ejection event of April 6, 2000: A practical demonstration of space weather nowcasting with the Super Dual Auroral Radar Network HF radars, *Journal of Geophysical Research*, *106*, 30085-30097, 2001.

Ruohoniemi, J.M., S.G. Shepherd, and R.A. Greenwald, The response of the high-latitude ionosphere to IMF variations, *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*, 159-171, 2002.

Ruohoniemi, J. M., and R. A. Greenwald, Dependencies of high-latitude plasma convection: Consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns, *J. Geophys. Res.*, *110*, doi: 10.1029/2004JA010815, 2005.

Schunk, R.W., L. Scherliess, J.J. Sojka, D.C. Thompson, D.N. Anderson, M. Codrescu, C. Minter, T.J. Fuller-Rowell, R.A. Heelis, M. Hairston, and B.M. Howe, Global assimilation of ionospheric measurements (GAIM), *Radio Sci.*, *39*, doi:10.1029/2002RS002794, 2004.

Schunk, R.W., and J.J. Sojka, A theoretical study of the lifetime and transport of large ionospheric density structures, *Journal of Geophysical Research*, *92*, 12343-12351, 1987.

Shepherd, S.G., J.M. Ruohoniemi, and R.A. Greenwald, Direct measurements of the ionospheric convection variability near the cusp/throat, *Geophysical Research Letters*, *30*, 2109, doi:10.1029/2003GL017668, 2003.

Shepherd, S.G., and J.M. Ruohoniemi, Electrostatic potential patterns in the high-latitude ionosphere constrained by SuperDARN measurements, *J. Geophys. Res.*, *105*, 23005-23014, 2000.

Valladares, C.E., S. Basu, J. Buchau, and E. Friis-Christensen, Experimental evidence for the formation and entry of patches into the polar cap, *Radio Science*, *29*, 167-194, 1994.

Weimer, D.R., Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients, *Journal of Geophysical Research*, *100*, 19595-19607, 1995.

1.3.4   Substorms

(Gordon Rostoker, Department of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2J1
rostoker@space.ualberta.ca)

Abstract

In the science of space weather, one of the most important disturbances under study is the magnetospheric substorm. In this review, the historical development of the study of substorms will be traced, the terminology used to describe the phenomenon will be introduced, and methods of modeling substorms will be presented. While the various frameworks of describing the substorm will not be reviewed here, the underlying physics of the process will be discussed and constraints limiting one's ability to interpret the existing data will be stressed.

1.3.4.1   Early Views of the Substorm

The term substorm was introduced by Akasofu [1] to describe a series of auroral activations observed using allsky cameras located at and around Fairbanks, Alaska. The activations followed a repeatable pattern and the terms used to describe this pattern are in use to this day. Figure 1-16 shows the sequence of events in what Akasofu termed an *auroral substorm*. The development of the substorm was initiated by a brightening of an auroral arc somewhere near the equatorward region of the region of auroras as detected by the allsky

cameras of the day. The initial brightening was followed by a rapid poleward evolution of the disturbed region, with the maximum poleward displacement being reached over a period of time averaging ~ 30 minutes. The period of time during which this poleward expansion took place was termed the *expansion phase* of the substorm. During the expansion *phase*, the western edge of the disturbed region developed a characteristic S-shaped form and was said to expand westward at an average speed of ~ 1km/s: this expanding auroral form was termed the *westward traveling surge*, and was deemed to move westward as far as the dusk meridian and beyond. After some minutes, the auroral arcs in the disturbed region began to drift equatorward, doing so until the original pre-substorm auroral configuration was reached after some tens of minutes. This interval of equatorward drift of auroral arcs was termed the *recovery phase*.
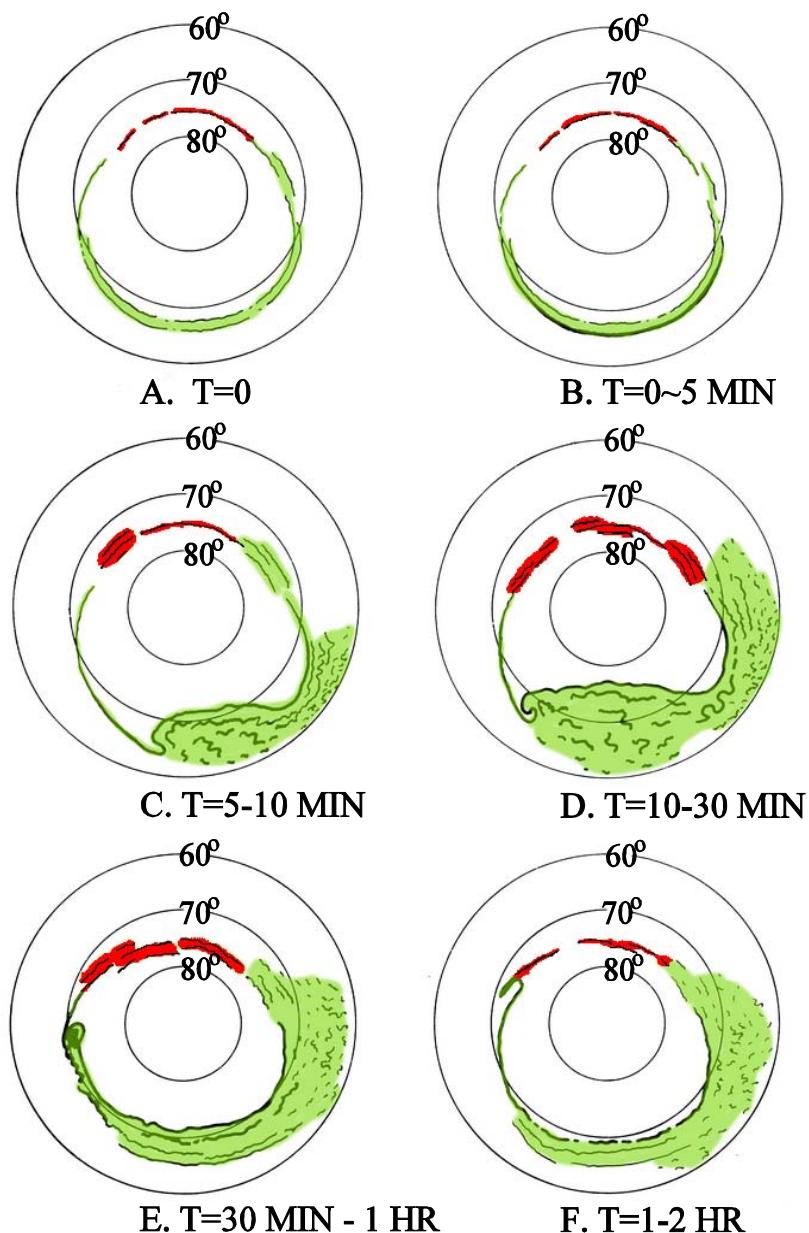
Figure 1-16: Sequence of Development of Discrete Auroral Arcs that Forms the Basis for the Substorm Expansion Phase Phenomenon (modified after Akasofu [1]). Since 1964, information has come to light about the diffuse aurora (green shading) in which the discrete auroral arcs are immersed and the dayside auroras (red shading) which tend to be enriched in red emissions.

It was well recognized that the substorm was accompanied by a marked intensification of electric current in the region of the ionosphere where the substorm arcs were observed. This increase in electric current produced a clear disturbance on a magnetogram (cf. Figure 1-17) which was called a magnetic bay (because its appearance on a magnetogram was similar to a bay on the coastline of a continent). Such disturbances had also been noted by researchers as far back as the late 19th century, and had been given the name *polar elementary storm* by Birkeland [2]. In fact, the auroral substorm was accompanied by many disturbances across the electromagnetic spectrum and Akasofu [3] ultimately brought together all these disturbances under the umbrella of the *magnetospheric substorm*. Within that umbrella, the low frequency magnetic disturbance was re-labeled the *polar magnetic substorm*, and most studies of substorms to this day have relied on the auroral and magnetic signatures of the substorm to define its presence or absence.
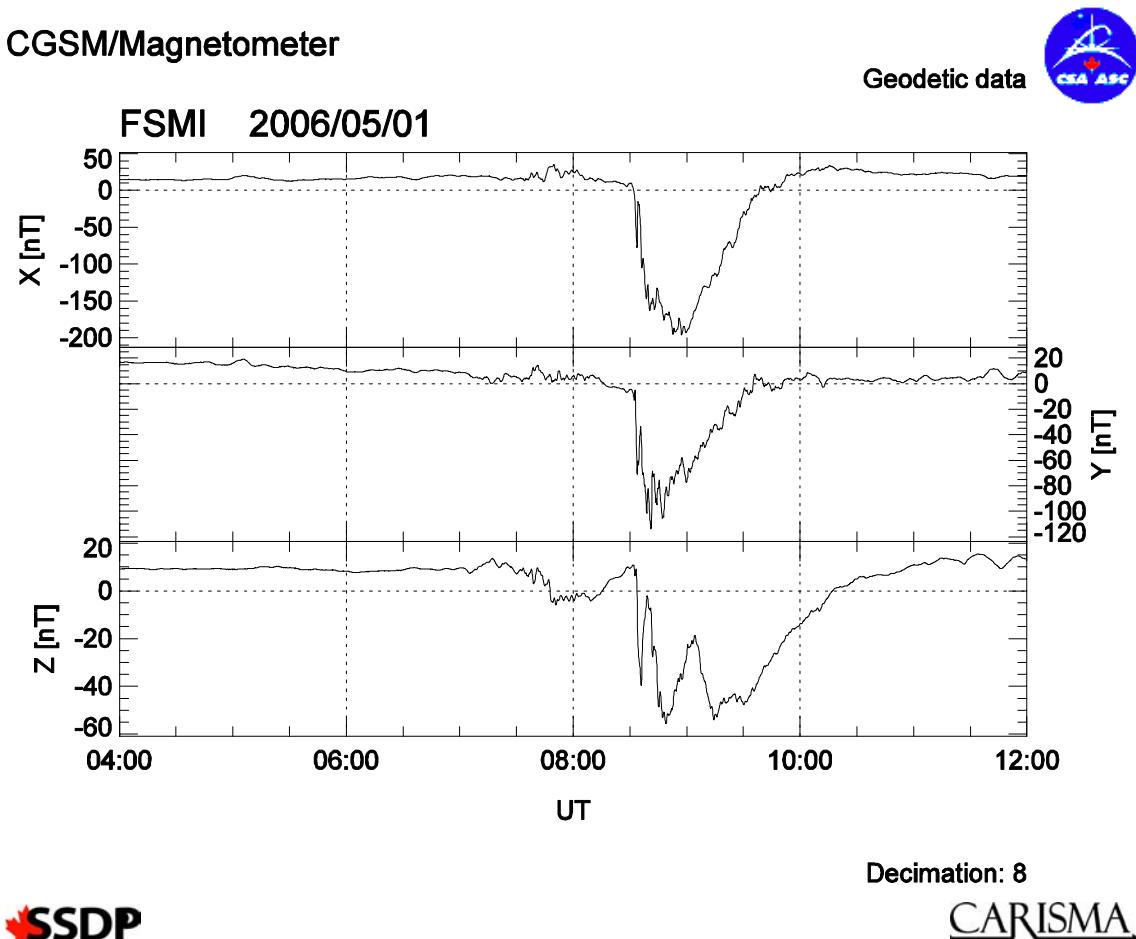
Figure 1-17: Typical Auroral Zone Magnetogram from Fort Smith, NWT Canada Showing the Development and Subsequent Decay of a Polar Magnetic Substorm. The shape of the time series of the negative X-component disturbance between ~083-0930 UT resembling a bay on a coastline gave rise to the term "geomagnetic bay" (data courtesy the Canadian Space Agency).

Based on the observations described above, researchers developed a paradigm for substorm physics at the beginning of the 1960's that involved energy transfer from the solar wind to the magnetosphere through the process of magnetic field line reconnection (cf. Dungey [4]). In this process, solar wind magnetic field lines merged at the front side magnetopause with terrestrial magnetic field lines. The merged field lines were swept over the poles to form a long stretched region of magnetic field behind the earth (i.e. the magnetotail) where they ultimately reconnected , sending the plasma on those field lines toward and away

from the earth on either side of the region of reconnection (i.e. the X-line or neutral line). The plasma that moved earthward became part of a large scale circulation pattern (cf. Axford and Hines [5]) in which plasma was transported into the near-Earth region while simultaneously being energized. In this process, the energy provided to the magnetosphere from the solar wind was often out of balance with the energy dissipation in the magnetosphere-ionosphere system. When the rate of energy input exceeded the rate at which energy could be dissipated, the excess energy was stored in the magnetotail magnetic field (i.e. the tail lobes). McPherron [6] ultimately proposed that there interval of time during which energy was stored in the tail lobes should be termed the *growth phase* of the substorm. From time to time, the stored energy would be suddenly deposited into the high latitude ionosphere, and these episodes of energy dissipation were thought to be associated with the substorm expansion phase through the engagement of the tail lobes and the reduction of the magnetic field energy stored in that region (cf. Atkinson [7]). This would be seen, in the ionosphere, as a decrease in the area of the polar cap working under the assumption that the (open) tail lobe field lines mapped to the polar cap. The reader is referred to Baker et al. [8] for a full review of the reconnection paradigm and the observations behind it, and to Kennel [9] for a review of the theoretical concepts behind the paradigm.

The relationship between the auroral and magnetic activity during intervals of substorm activity was the subject of strong controversy over the early years after the term *substorm* was introduced. Prior to the introduction of the term, researchers had studied the phenomenon though its magnetic signatures (called geomagnetic bays) that accompanied the intervals of enhanced auroral activity. Two characteristic types of magnetic disturbance were noted, as characterized by the equivalent current systems that were determined from available ground magnetometers distributed around the world. [An equivalent current system is drawn under the assumption that all magnetic perturbations are due to sheet overhead current flow. The direction of the equivalent current is at 90° to the direction of the horizontal magnetic field perturbation at each measurement site and the separation of the equivalent current lines is inversely proportional to the strength of the magnetic disturbance.] The two systems featured two cells and one cell of equivalent current (cf. Figure 1-18) and have been given different names by different researchers over the years. The two cell system has been called SD (Chapman and Bartels [10]), DP2 (Obayashi [11]) and the directly driven current system (Akasofu [12]). The one cell system has been called DP1 (Obayashi [11]), the unloading component of the substorm (Akasofu [12]) and the substorm current wedge (Baumjohann [13]). Fukushima [1953] had noted that sometimes geomagnetic activity was best represented by a two cell system and sometimes by a one cell system. After the introduction of the term substorm by Akasofu [1], Sugiura and Heppner [14] suggested that this disturbance was best characterized by the two cell system while Akasofu et al. [15] suggested that the substorm should be associated with the one cell system. Ultimately, Rostoker [16] demonstrated that both systems coexisted during substorm activity, setting the stage for the generally held view at present that a substorm involves both directly driven activity an a process for storage and release of energy from the magnetotail, the release being responsible for the substorm expansion phase. Figure 1-19 shows a cartoon of the two components of substorm activity as reflected in the auroral electrojets and the field-aligned currents that connect those electrojets to the source regions of the outer magnetosphere.
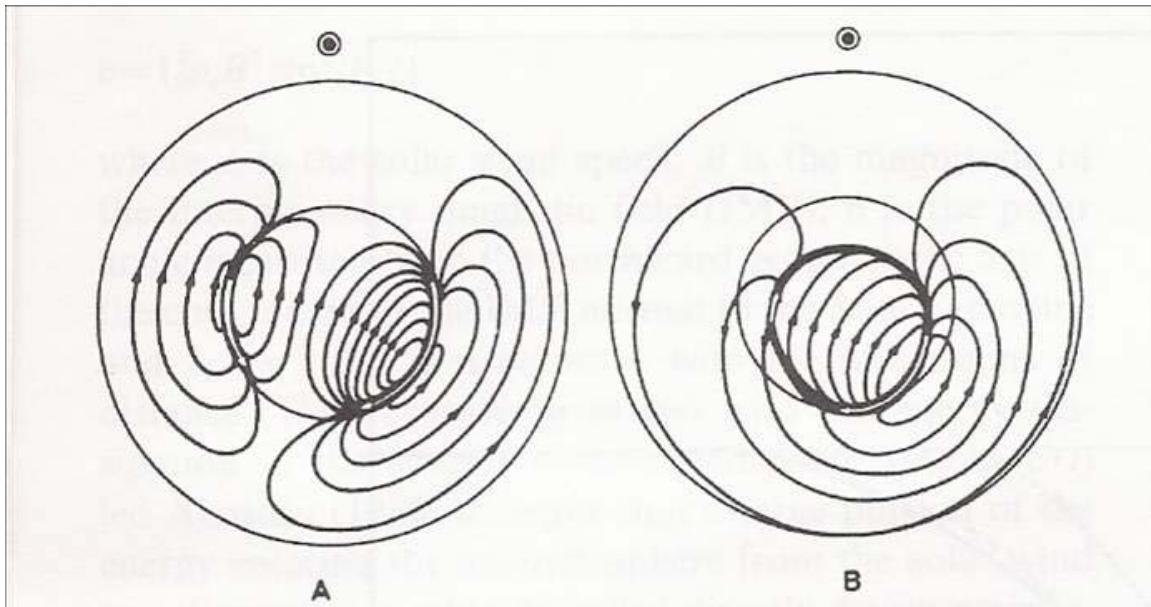
Figure 1-18: The Two Cell and One Cell Equivalent Current Systems which Both Had, at One Time or Another, Been Considered to be Representative of the Substorm Disturbance (after Rostoker [29]).
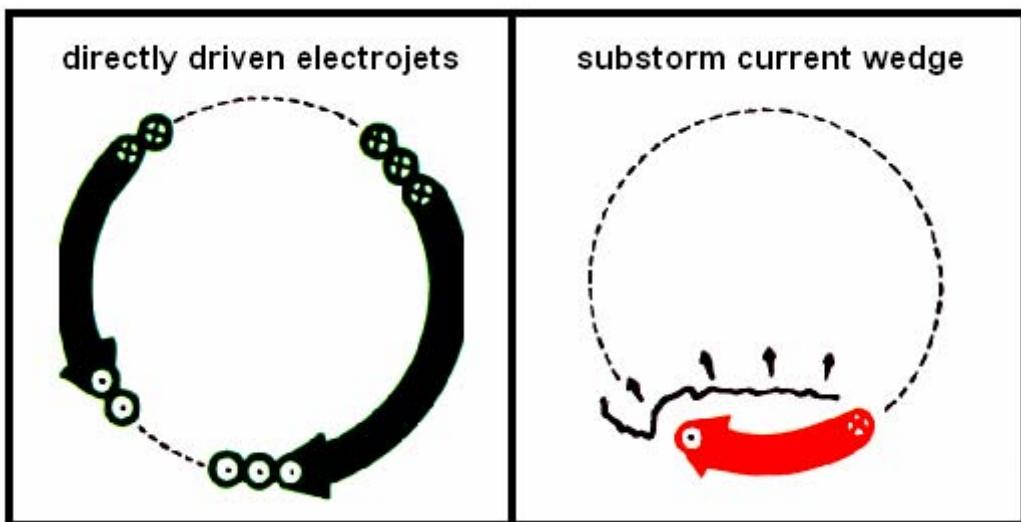
Figure 1-19: Ionospheric and Field-Aligned Currents Associated with the Auroral Electrojets for the Two Components of the Substorm Disturbance (modified after Baumjohann [13]).

1.3.4.2   Present View of the Substorm

It should come as no surprise that the original template for a substorm constructed by Akasofu [1] has undergone considerable evolution. The measurement tools of the time (i.e. allsky cameras) suffered from low film sensitivity and a very sparse distribution of the earth's surface. They were unable to distinguish the presence of diffuse aurora and hence fully define the latitudinal range of energetic particle precipitation into the auroral ionosphere. Vastly improved instrumentation including global imaging from orbiting satellites has afforded researchers a much more detailed picture of the evolution of a substorm, and theoretical studies have provided a framework in which many substorm. observations can be understood in the context of solar wind-magnetosphere coupling and magnetosphere-ionosphere coupling. In particular, knowledge of how

substorm activity develops in response to changes in solar wind energy input has provided the opportunity to better place the Akasofu [1] substorm development in a larger global context.

The understanding of the physical processes which lead to substorm activity requires that the evolution of the disturbance be placed in the framework of changes in the source of energy, viz. the solar wind. As mentioned in the previous section, the development of substorm activity is associated with an increase in the input to the magnetosphere of solar wind energy. This increase is most often affected by the IMF becoming more southward, increasing the rate of reconnection at the front side magnetopause. The increased energy is partially dissipated in the ionosphere through Joule heating and particle precipitation associated with the directly driven currents (c.f. left panel of Figure 1-19) with the balance of the energy being stored in the magnetotail magnetic field. Figure 1-20 shows the development of activity in the auroral oval in response to the increase in energy input from the solar wind. After the southward turning of the IMF, the size of the polar cap grows and the auroral region is forced to lower latitudes. During this growth phase of the substorm, the inner edge of the plasma sheet moves earthward and the plasma sheet, itself, is observed to become thinner. Ultimately, at some later time, the IMF returns to its previous configuration, often by a sudden northward turning. This change in external conditions often triggers the substorm expansion phase, which initially involves the brightening of auroral arcs near the equatorward edge of the midnight sector auroral oval.
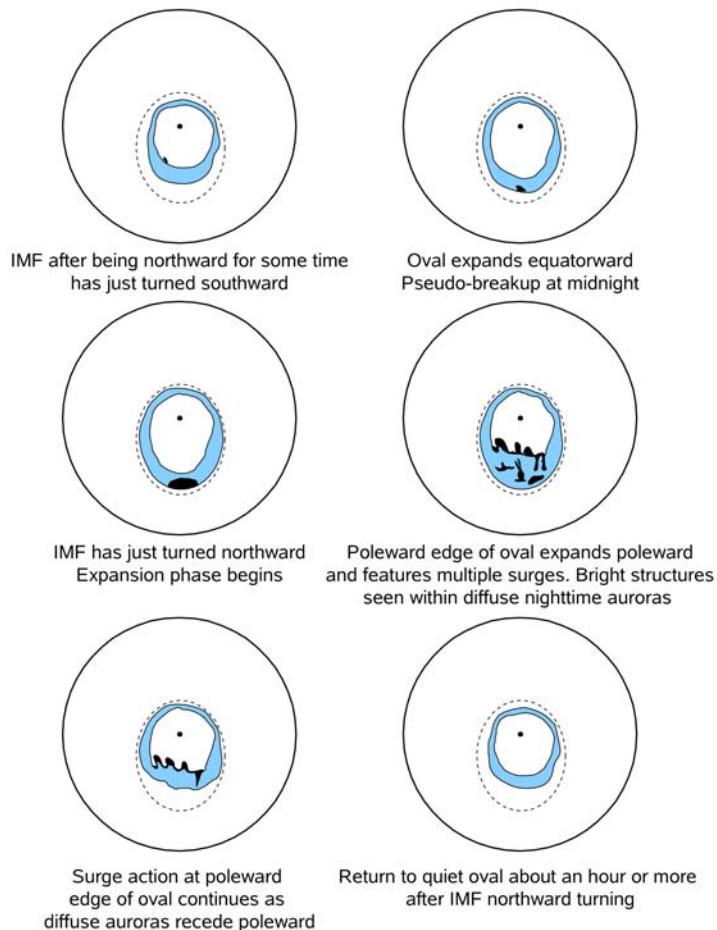
Figure 1-20: A Global Perspective of the Behaviour of the Auroral Oval During the Growth and Expansion Phases of a Substorm. The Akasofu [1] auroral substorm development characterizes the behaviour of the auroras around midnight after a reduction in energy flow into the magnetosphere (often triggered by a northward turning of the IMF), featuring an intensification near the equatorward edge of the oval and subsequent explosive poleward motion of the high latitude edge of the disturbed region.

The explosive poleward motion of the breakup region is well described by the original Akasofu framework shown in Figure 1-16. In a matter of less than 5 minutes, the disturbed region has spread to the poleward edge of the oval, and from then on the polar cap collapses back to its pre-substorm configuration. During the minutes to tens of minutes over which time the polar cap collapse takes place, bright auroral loops and surges appear along the poleward edge of the evening sector oval. These disturbances have been termed *poleward boundary intensifications* (PBIs) and their auroral luminosities and associated magnetic perturbations are commensurate with those detected during the explosive development of the substorm disturbed region at onset.

One of the most important realizations coming from satellite borne imagers is that there are two distinct regions of auroral activations in the night side auroral oval. One region is located near the equatorward edge of the midnight sector auroral oval, and is the site of substorm expansive phase onset and pseudo-breakups (which have the appearance of an expansive phase onset but engage only the closed field line portion of the magnetotail and hence never develop to the extent of causing a reduction in the area of the polar cap). Figure 1-21 shows the two zones of the double oval, from which it is seen that the regions of auroral luminosity in the midnight/morning and evening sectors are the sites of unbalanced upward field-aligned current (FAC). These two zones can both be activated and may feature auroral loops and surges whose magnetic and auroral signatures are quite similar. Figure 1-22 shows a moment during an interval of strong magnetospheric activity. In Panel a of that Figure, the two branches of the evening sector are clearly evident and an auroral surge form is seen on the most poleward of the two branches. This surge form is a PBI, and may occur either with or without the presence of substorm expansive phase activity. In Panel b of the Figure, a substorm expansion phase is seen to be initiated in the late evening sector of the equatorward branch of the double oval.
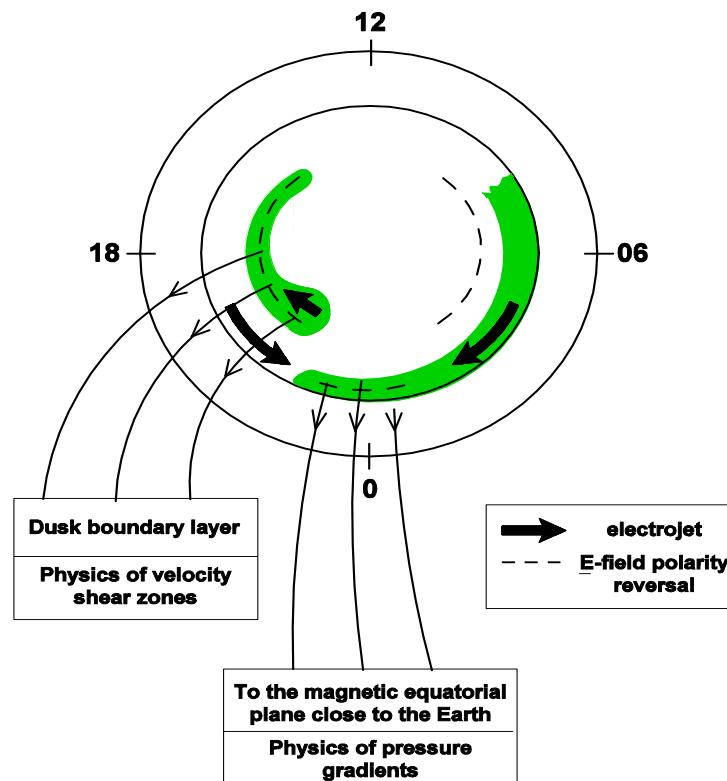
Figure 1-21: The Two Zones of Auroral Luminosity (shown in green) Associated with the Double Oval. These regions of discrete auroral arcs mark the presence of upward field-aligned current that connects the ionosphere to the source regions in the magnetosphere where energy is provided. The eastward and westward electrojets are bounded by anti-parallel Birkeland current sheets (cf. Zmuda and Armstrong [41]) not shown in this figure.
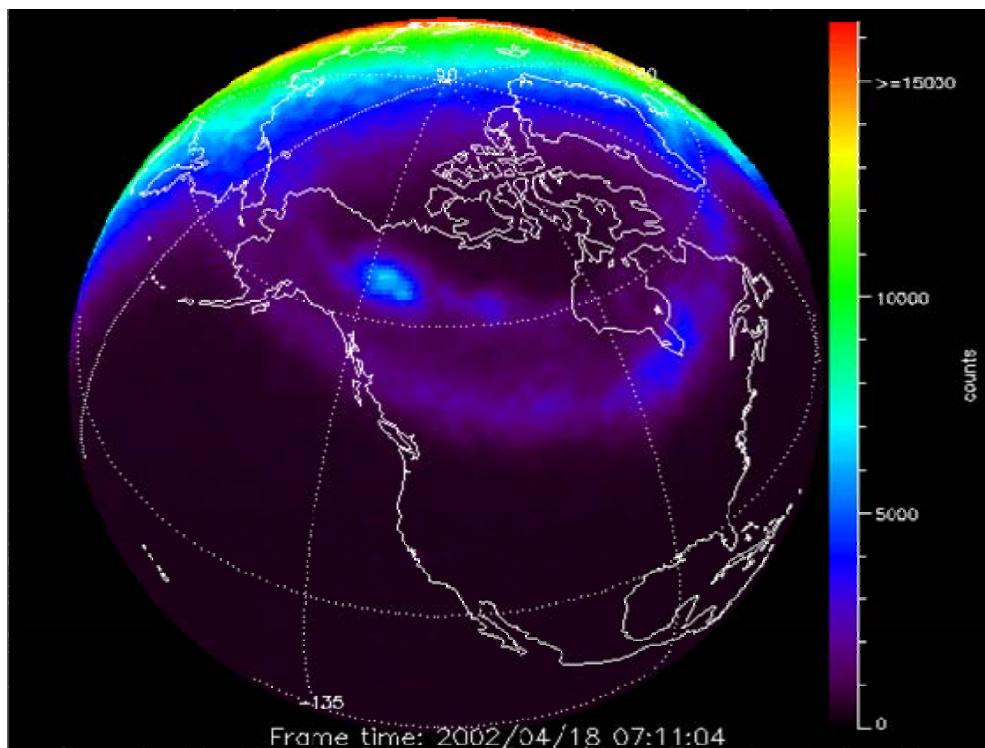
Figure 1-22a: A Poleward Boundary Intensification (PBI) on the Poleward Branch of the Double Oval as Observed by the UV Camera Aboard the IMAGE Satellite (courtesy S. Mende [University of California, Berkeley] and CDAWeb).
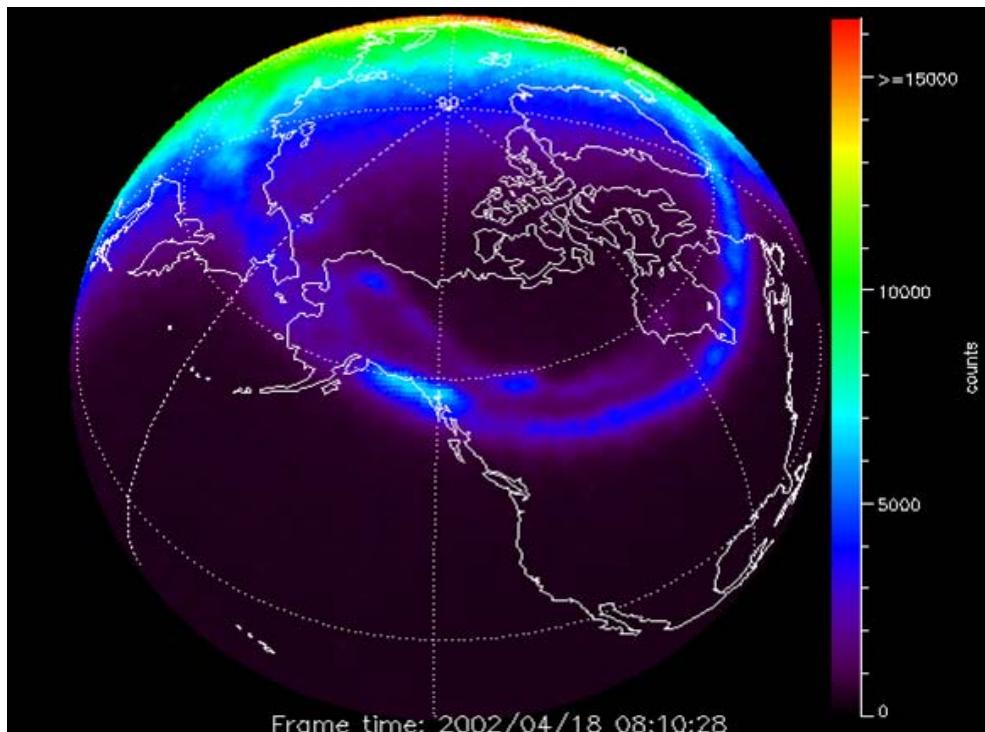
Figure 1-22b: A Substorm Expansion Phase Onset on the Equatorward Branch of the Late Evening Sector Auroral Oval as Observed by the UV Camera Aboard the IMAGE Satellite (courtesy S. Mende [University of California, Berkeley] and CDAWeb).

Finally, it should be noted that one of the most spectacular auroral features, the so-called westward traveling surge (WTS), may not travel much at all in the sense of the moment of the auroral luminosity shifting westward over significant distances. The concept of the WTS originated from observations by allsky cameras of the local field of view. The surge appeared as a bright luminous form at the eastern edge of the field of view and its western edge would shift westward across the field of view at speeds of ~ 1 km/s on the average. As early as the 1970's, Kisabeth and Rostoker [17] noted that, while the western edge of the surge form expanded westward, the eastern edge might expand eastward. This behavior is characteristic of a spatial confined region of luminosity that grows by expanding in all directions from its initial site of formation. Subsequently, Tighe and Rostoker [18] showed that the initial westward expansion of the western edge of a surge form could occur at speeds of several km/s in the early stages of development, but that the expansion would cease after a few minutes leaving behind a quasi-stationary bright auroral form. There could, in fact, be several such forms arrayed along the high latitude boundary of the evening sector auroral oval during the course of a substorm event. The results of Craven et al. [19] support this revised view of the WTS, and it would appear that the concept of a monolithic surge form moving over large longitudinal distances during the course of a substorm is no longer tenable. It does appear that the surge forms are actually poleward boundary intensifications (PBIs) in the light of their location at the poleward edge of the disturbed auroral oval after the poleward expansion of the auroral bulge during the substorm expansion phase.

1.3.4.3   Storm-Substorm Relationship

Much of the early work on the solar-terrestrial interaction involved the study of the response of the ionosphere to geomagnetic storms. These storms featured a significant decrease in the horizontal component of the low latitude magnetic field that grew in a matter of hours and decayed over a period of tens of hours to a few days. These disturbances, ranging magnitude from a few tens of nT to hundreds of nT were understood as the consequence of the formation of a symmetric ring current flowing around the earth at distances of ~ 3-8 Re. The strength of this ring current was quantified through the Dst index (cf Sugiura [20]), which was computed from the horizontal (H) component of the magnetic field from ~ 4 – 5 low latitude stations distributed with approximately equal longitudinal separation around the world.

While it was clear from the beginning (cf. Rostoker and Fälthammar [21]) that southward interplanetary field was required for the development of a geomagnetic storm, the role of substorms in that development has been somewhat controversial. Akasofu et al. [22] argued that substorms played an integral role in the formation of the storm time ring current by injecting energetic particles from the plasma sheet to regions much closer to the earth where they ultimately became the ring current carriers. However, shortly thereafter, Burton et al. [23] provided convincing evidence that simply the presence of a sufficiently strong dawn-to-dusk interplanetary electric field for a sufficiently long period of time was sufficient to produce the observed ring current growth. In the scheme of Burton et al., there was no role for substorms in the ring current development, although it was clear that substorms and the ring current derived their energy from the same source, i.e. the solar wind.

Since that time, the role of substorms in the creation of the storm time ring current has been a subject of intense research activity. This question has been highlighted in reviews of the putative storm-substorm relationship that have appeared in recent years (cf. Gonzalez et al. [24]; Kamide et al.[25]; Sharma et al.[26]) but no definitive assessment had emerged. However, Tsurutani et al. [27] took advantage of the topside imaging capability of the Polar spacecraft to provide evidence that, during the main phase of a magnetic storm, there could be no substorm activity although many spatially confined auroral structures (e.g. pseudo-breakups, torches, poleward boundary intensifications) were in evidence. Zhou et al. [28] explored this question further and reached the same conclusion, namely substorms were not an essential element of storm time ring current growth.

On the other hand, Rostoker [29] has argued that the process whereby substorm expansion phase is initiated near the inner edge of the plasma sheet plays an integral role in the development of the storm time

ring current. He contends that pseudo-breakups and substorm expansion phase onsets both reflect the results of an instability near the inner edge of the crosstail current sheet which initiates the breakdown of the shielding electric field (cf. Senior and Blanc [30]). This breakdown permits the penetration of the convection electric field closer to the earth, leading to additional acceleration of plasma sheet particles and increasing the possibility that they can be injected into stable trapped orbits as part of the ring current population. For a classical substorm, triggered by a reduction of energy input into the magnetosphere from the solar wind (cf. Rostoker [31]), the effects of the earthward penetration of the convection electric field is mitigated by the fact that that electric field begins to decline in strength. For storm conditions, when the driving electric field maintains its strength for lengthy periods of time, there can be a series of breakdowns of shielding, each episode occurring closer and closer to the earth. In this way, the inner edge of the plasma sheet can get relatively close to the earth and the probability increases that injected energetic particles can become part of the ring current population.

1.3.4.4   Conductivity of the Auroral Ionosphere

One of the primary mechanisms for dissipation of energy entering the magnetosphere from the solar wind is Joule heating associated with the flow of electric current through the resistive ionosphere. The entire process starts with the penetration of solar wind plasma into the outer edges of the magnetosphere. The kinetic energy of flow of this plasma is converted to electromagnetic energy, with a major portion of this being stored in the magnetic field of the tail lobes and in the magnetic field associated with the large scale current systems that couple the source region of the outer magnetosphere to the auroral ionosphere.

The Joule heating takes place in a relatively limited region of altitude extending from ~ 100 – 300 km above the earth's surface. Here, there is a significant electric conductivity as a consequence of collisions between the neutral particles of the atmosphere and the ions and electrons that are there as a consequence of collisions with energetic particles precipitating from the magnetosphere and with the interaction of solar visible and UV radiation with the atmospheric neutrals. The precise number densities of ions and electrons represent a balance between the production rates due to particle precipitation and solar radiation and recombination rates due to collisions among the atmospheric constituent particles.

The ionospheric conductivity at a point is space is anisotropic, and is typically expressed in terms of current flow parallel to the component of the electric field normal to the ambient magnetic field and perpendicular to that component of the electric field. These are termed the Pedersen and Hall conductivities respectively, and are defined by the expressions:

$$\sigma_P = \frac{n_e}{B}\left[\frac{\nu_e \omega_e}{\omega_e^2 + \nu_e^2} + \frac{\nu_i \omega_i}{\omega_i^2 + \nu_i^2}\right]$$

$$\sigma_H = \frac{n_e}{B}\left[\frac{\omega_e^2}{\omega_e^2 + \nu_e^2} - \frac{\omega_i^2}{\omega_i^2 + \nu_i^2}\right]$$

where n_e is the number density of electrons, B is the magnetic field, ν_e and ν_i are the electron-neutral and ion-neutral collision frequencies, and ω_e and ω_i are the electron and ion gyrofrequencies. There is, of course, a conductivity parallel to the magnetic field, but this is generally assumed to be extremely high consistent with the assumption in magnetohydrodynamics of frozen field conditions over most of the length of the magnetic field lines linking the ionosphere to the outer magnetosphere.

Often, one sees, rather than the conductivity at a point, the height integrated conductivities:

$$\Sigma_P = \int \sigma_P dz$$

$$\Sigma_H = \int \sigma_H dz$$

where the integration typically extends from the bottom of the ionosphere (z~ 90 km) to the top of the ionosphere (z ~ 300 km). All the conductivities measured observationally or used in theoretical expressions describing magnetosphere-ionosphere coupling are height-integrated and it is generally assumed that the current in the ionosphere flows in an infinitesimally thin sheet for which the current flow is described by:

$$\mathbf{J} = \Sigma_P \, \mathbf{E} + \Sigma_H \, \mathbf{n}_r \times \mathbf{E}$$

where $\mathbf{n}_r$ is a unit radial vector, $\mathbf{E}$ is the component of the electric field perpendicular to $\mathbf{B}$ and frozen-field conditions apply, viz. $\mathbf{V} = (\mathbf{E} \times \mathbf{B})/B^2$. From the assumption that div $\mathbf{J} = 0$, one can obtain an expression for the field-aligned current density:

$$j_{\parallel} = \Sigma_P \, \text{div} \, \mathbf{E} + \mathbf{E} \cdot \text{grad} \, \Sigma_P + \mathbf{n}_r \times \mathbf{E} \cdot \text{grad} \, \Sigma_H$$

where div $\mathbf{V} = 0$. From this expression, it can be seen that knowledge of gradients in the Hall and Pedersen conductivities are crucial in determining where field-aligned currents flow. Under conditions of steady magnetospheric convection, the location of the discrete aurora in Figure 1-21 follows automatically from consideration of the Pedersen conductivity terms in Eq (4) when one understands that the discrete aurora mark the location of upward field-aligned current.

A number of researchers have produced semi-empirical models of ionospheric conductivity that are available for use in simulations of electric current flow in the magnetosphere-ionosphere system. One of the earliest of these models was produced by Wallis and Budzinski [32], who used measurements of precipitating energetic electron fluxes from the Isis 2 satellite together with the theoretical estimates of ionization rate per unit volume formulated by Rees [33, 34] to obtain estimates of the Hall and Pedersen conductivities every two hours of magnetic local time at one degree intervals of geomagnetic latitude from 60-84°. They included in their calculations the conductivity effected by solar radiation and considered conditions of relatively low activity (0 < Kp < 3) and of high activity (3 < Kp < 9). Since the work of Wallis and Budzinski, there have been several other efforts to estimate ionospheric conductivity, most of them calling on measurements of precipitating electron flux (e.g. Spiro et al.[35]; Robinson et al. [36]; Hardy et al. [37]).

For example, Robinson et al.[36] developed semi-empirical expressions relating the height-integrated Hall and Pedersen conductivities to the average energy and energy flux of the precipitating electrons responsible for the conductivity, i.e.

$$\Sigma_P = \frac{40\overline{E}}{16 + \overline{E}^2} \Phi_E^{1/2}$$

$$\frac{\Sigma_H}{\Sigma_P} = 0.45\overline{E}^{0.85}$$

A different approach was developed by Ahn et al. [38] who developed semi-empirical relationships between the height-integrated Hall and Pedersen conductivities and the magnetic perturbation $|\Delta H|$ in the north-south component of the magnetic field, viz.

$$\Sigma_{H}, \Sigma_{P} = a \mid \Delta H \mid^{b}$$

that were parameterized to apply to the poleward and equatorward portions of the auroral oval for the evening (eastward electrojet) and morning (westward electrojet) sectors.

The question now arises as to how useful these statistical procedures are in addressing how substorms influence ionospheric conductivity and hence the flow of electric current between the generator regions in the outer magnetosphere and the sink of energy in the upper atmosphere. A major source of difficulty is that all relationships were developed based on the statistical properties of the individual parameters in terms of their variation with activity level as quantified by an index (e.g. Kp) or by individual site measurements (e.g. perturbations in the north-south component of the magnetic field $|\Delta H|$ measured on the ground). Because the currents in the auroral oval can produce similar magnetic perturbations for different latitudinal positioning of the auroral oval, there is certain to be a smearing for any parameter inferred from such data. The same is true for the flux and average energy of precipitating electrons for a given activity level. Figure 1-23 shows Hall conductivities inferred from precipitating energetic electron fluxes by Hardy et al. [39] for varying levels of magnetic activity as quantified by the Kp index. One would expect to see in the evening sector two zones of enhanced conductivity, consistent with the distribution of upward field-aligned current shown in Figure 1-21. For the Kp = 2 pattern, this is in fact observed, but for the Kp = 3 pattern smearing of the data for both branches of the double oval makes the distribution appear more continuous. In reality, as one moves to higher activity levels, the probability of the data being taken during an substorm expansion phase is enhanced, and for such circumstances one would expect the midnight and late evening auroral oval to feature many spatially localized regions of high luminosity, varying rapidly in space and time. Any statistical study of energetic electron precipitation binned according to latitude is very likely to produce a smeared out latitudinally confined region of enhanced fluxes, even though individual events may actually feature discrete structures such as the double oval or longitudinally confined north-south streamers (cf. Rostoker et al. [40]).
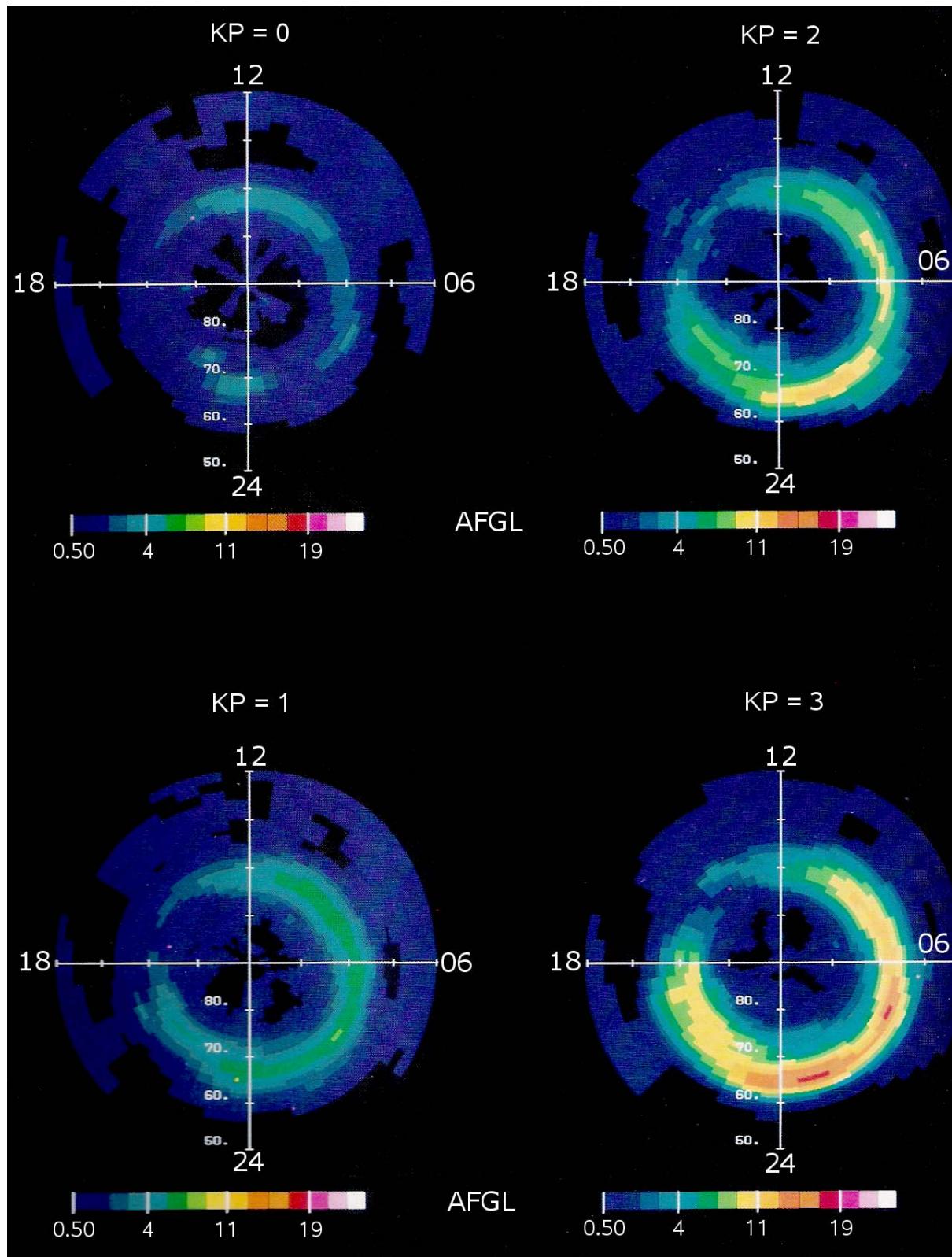
Figure 1-23: The High Latitude Hall Conductivity Distribution Inferred from Precipitating Energetic Electron Fluxes at Different Magnetic Activity Levels as Quantified by the Kp Index (Top left Kp=0, Bottom left Kp=1, Top right Kp=2 , Bottom right Kp=3). The colors are assigned on a logarithmic scale (after Hardy et al. [37]).

1.3.4.5   Conductivity in the Ionosphere During Substorm Activity

In section 1.2 above, the general conductivity distribution during periods of magnetospheric activity has been presented. Technically, one should try to decouple the conductivity distribution during times of purely directly driven activity (i.e. steady magnetospheric convection) from that in evidence during substorm expansion phase activity. However, this has never been done because of the difficulty in identifying periods of purely directly driven activity. There are, however, two aspects of substorm activity in which specific conductivity situations have been identified and discussed in the context of coincident electric field changes.

The first of these relates to the onset of the substorm expansion phase, at which time breakup arcs appear near the equatorward edge of the midnight sector auroral oval. At this time, it is suggested that the westward convection electric field penetrates closer to earth as a consequence of a reduction in the shielding electric field (cf. Rostoker [29]). A consequence of this is the development of a southward polarization electric field across the region of auroral breakup, so that the total horizontal electric field in the midnight sector auroral ionosphere points in the south-west direction. When the orientation of this electric field is such that the southward Pedersen current exactly cancels the northward Hall current, the intensity of the westward electrojet is a maximum. Under these circumstances, the westward electrojet is called a Cowling current and its current density is written

$$J_C = \Sigma_C \, E_w$$

where $\Sigma_C = (\Sigma_H^2 / \Sigma_P) + \Sigma_P$ is the Cowling conductivity.

The second issue, somewhat related to the first, is the electric current flow in the vicinity of an auroral surge form. As mentioned earlier, auroral surges often turn out to be PBIs located at the poleward boundary of the evening sector auroral oval. An auroral surge form represents a concentrated region of precipitating energetic electrons, and hence a region of upward field-aligned current. The scale size of surges is ~ 500 km when fully developed, and the western edge of a surge can expand westward initially at speeds of ~ 5 km/s although, after a few minutes, the western edge comes to a halt (cf. Tighe and Rostoker [18]). A surge can be viewed as a core region of upward field-aligned current, encircled by an annular region of downward field-aligned current. The downward current connects to the upward current in the ionosphere through radially flowing Pedersen current, and in the magnetosphere with a generator current in the source region where the energy is available to accelerate particles into the ionosphere and to permit the degrading of stored energy in the magnetosphere to heat in the ionosphere though the process of Joule heating. There is also a Hall current that circulates around the luminous region of the surge, flowing equatorward at the western head of the surge and poleward at its eastern trailing edge. The magnetic effects of this vortical Hall current are what produce the characteristic signatures observed by magnetometers on the ground in the vicinity of the surge form. Figure 1-24 shows ground magnetometer data from two sites in northern Alaska which are fortuitously located at the western edge (Barrow) and the eastern edge (Katovik) of a large auroral surge form detected at ~0600 UT (aurora not shown here but observed by the UV imager aboard the Polar satellite). The positive D-component at Barrow is caused by equatorward ward flowing Hall current at the western edge of the surge, while the negative D-component at Katovik is caused by poleward flowing Hall current at its eastern edge. The positive H-component is consistent with the two stations being at the equatorward edge of the bright auroras associated with the surge.

1999/12/07
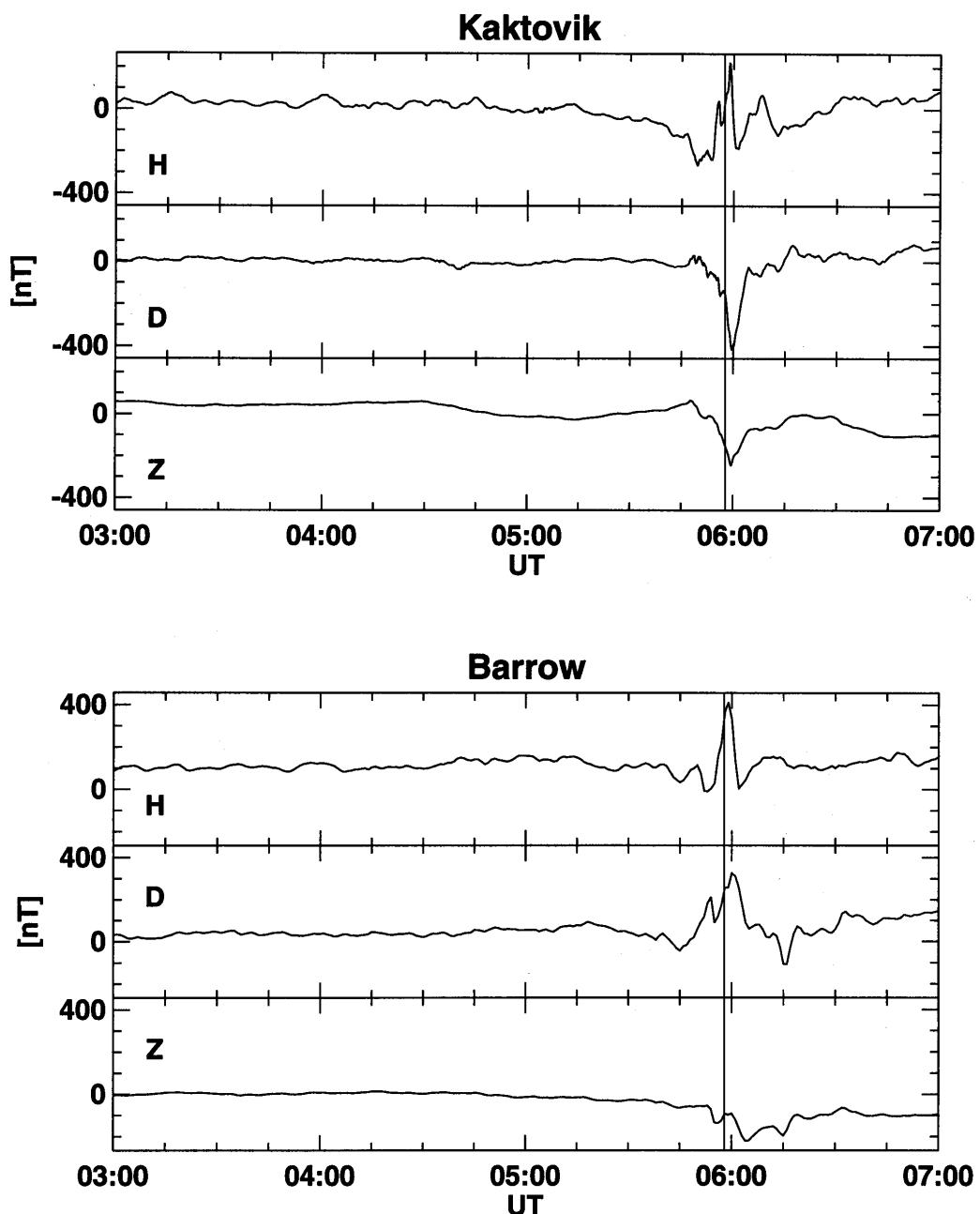
Kaktovik

Barrow

Figure 1-24: Ground Magnetometer Data from Two Stations in Northern Alaska (Barrow and Katovik), Located at Approximately the Same Magnetic Latitude but Separated in Longitude. Fortuitously, Katovik is situated on the eastern edge of an auroral surge while Barrow is situated on the western edge of the same surge.

Finally, one should note that the aurora during substorm expansive phase activity involve discrete arcs, often separated by regions of darkness. Therefore, the conductivity distribution can be expected to be very complex and images obtained from high altitude satellites some thousands of km above the aurora cannot be expected to provide information on gradients with a resolution of any better than ~100 km.
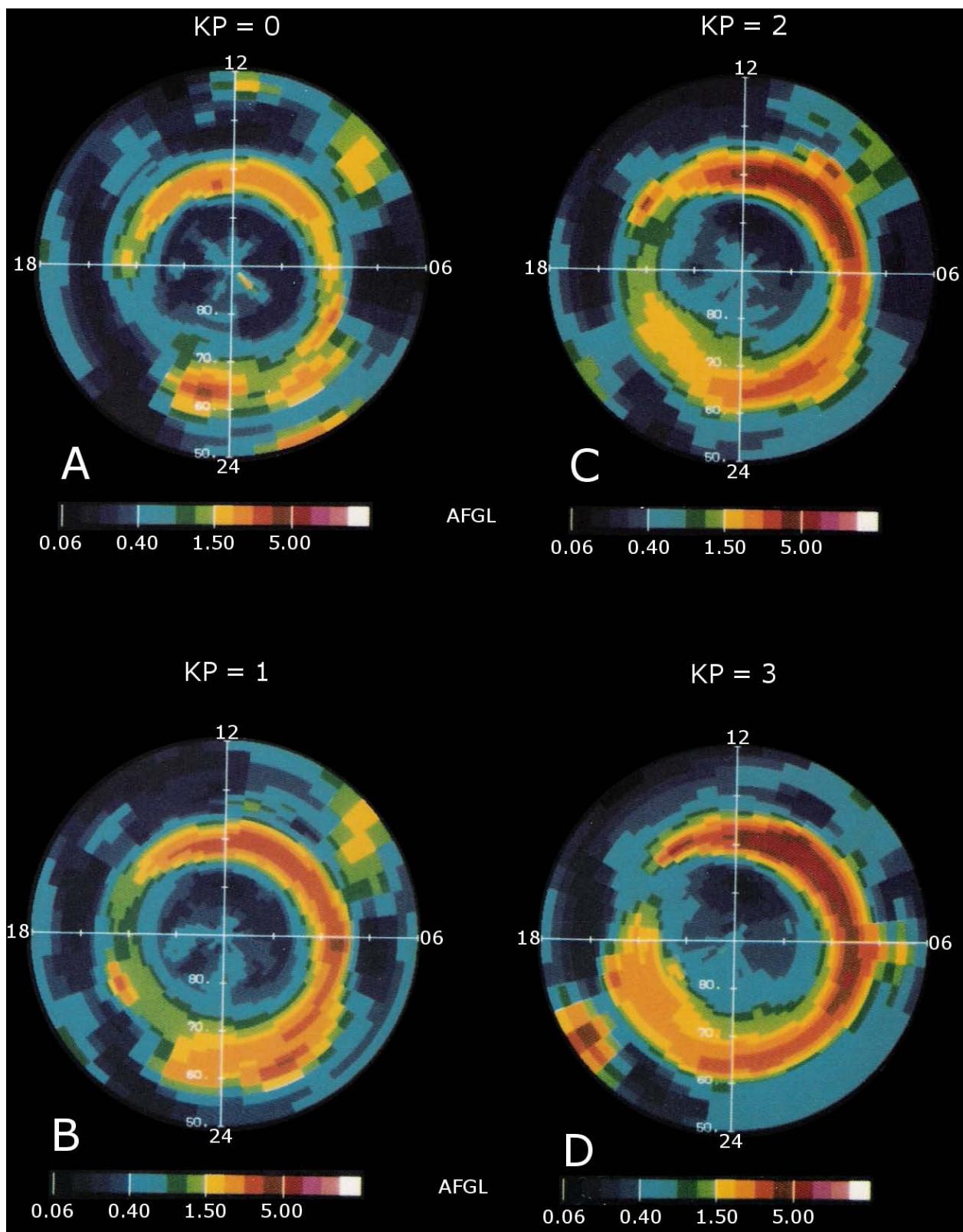
Figure 1-25: Polar Plots of the Average Energy of Precipitating Electrons, in Units of keV, for Different Activity Levels as Quantified by Kp (Panel A Kp=0, Panel B Kp=1, Panel C Kp=2, an Panel D Kp=3). The colors are assigned on a logarithmic scale (after Hardy et al. [39]).

1.3.4.6   Precipitation of Energetic Particles into the Ionosphere

For practical purposes, the primary way in which substorm activity influences the high latitude ionosphere is to alter the electrical conductivity in the auroral oval in a way that changes the ionospheric current flow and the distribution of field-aligned upward and downward currents that link the source regions of the outer magnetosphere to the dissipative upper atmosphere. To quantify that behaviour, efforts have been made in the past to establish the local time distribution of energetic electron precipitation for varying levels of magnetospheric activity. The most detailed data base analysed thus far was acquired using the DMSP satellites and published by Hardy et al. [39] who binned the data by activity level as defined by the Kp index. Examples of the global distribution of the average energy of precipitating electrons are shown in Figure 1-25 for four different levels of Kp ranging from quiet to disturbed conditions. For quieter conditions (panels A and B), it is apparent that the energetic electron precipitation maximizes at lower latitude in the midnight and post midnight sectors and at higher latitudes in the pre-midnight sector. This is consistent with the concept of the double oval mentioned earlier in this review, and with the observations of upward field-aligned current made by polar orbiting satellites carrying magnetometers (cf. Zmuda and Armstrong [41]) since strong upward field-aligned currents are carried by energetic electrons precipitating into the ionosphere.

It should be pointed out that the simple patterns shown in Figure 1-25 are what one would expect in the absence of substorm expansion phase activity. When expansion phase activity is in progress, it is often difficult to separate the poleward and equatorward branches in the double oval and statistical representations of precipitating energetic electron fluxes seem to indicate a smooth connection across midnight of the low latitude region of energetic electron precipitation in the morning sector with the high latitude region of precipitation in the evening sector. Great care has to be exercised in the study of particles and fields in the hours around midnight when substorm expansion phase activity is in progress.

1.3.4.7   Modeling of Substorm Current Systems

The product of several decades of experimental and theoretical studies of the solar-terrestrial interaction has been the development of models that synthesize the data from observation points in space and on the ground and produce global pictures of the disturbances associated with the transport and dissipation of energy derived from the solar wind. Since there is no long term stable configuration of satellites available for modeling studies, it has fallen on ground based data to provide of the core of information required to produce models of real electric currents and electric potentials which are required to quantify the character of the coupling between the magnetosphere and ionosphere which determines the dispensation of the solar wind energy that penetrates into the magnetosphere.

Until the late 1970's, the only way in which global patterns of energy dissipation were described in a quasi-quantitative fashion was through the use of equivalent current systems of the type shown in Figure 1-18. While these were useful, they did not take into account the fact that a significant portion of the horizontal component of the magnetic field disturbance could be due to field-aligned rather than ionospheric current flow. Techniques developed independently by Mishin et al. [42] and Kamide et al. [43] provided the ability to obtain, simply from ground based measurements, estimates of the real current flow (both ionosphere and field-aligned) and the cross-polar cap potential drop. Following the approach of Kamide et al. [43], the flow lines of the equivalent current system were characterized by a scalar potential function Ψ so that the equivalent current density (in A/m) is given by

$$\mathbf{J}_T = \mathbf{n}_r \times \text{grad } \Psi$$

The total height-integrated current is represented as the sum of the equivalent current and a potential current $\mathbf{J}_p$, i.e.

$$\mathbf{J} = \mathbf{J}_p + \mathbf{J}_T$$

where div $\mathbf{J}_T = 0$ and curl $\mathbf{J}_p = 0$. The field-aligned current density is then given by

$$j_{\parallel} = \text{div } \mathbf{J} = \text{div } \mathbf{J}_p$$

and $j_{\parallel}$ is positive downwards.

The real ionospheric current, expressed in the frame of reference co-rotating with the earth, is written as

$$\mathbf{J} = \Sigma_P \mathbf{E} + \Sigma_H \, \mathbf{E} \times \mathbf{n}_r$$

and the electric field is assumed to be electrostatic, i.e. $\mathbf{E} = $ -grad Ø.

Using the above equations, one can write a partial differential equation for Ø in terms of Ψ

$$A \frac{\partial^2 \emptyset}{\partial \theta^2} + B \frac{\partial \emptyset}{\partial \theta} + C \frac{\partial^2 \emptyset}{\partial \lambda^2} + D \frac{\partial \emptyset}{\partial \lambda} = F$$

where θ is the colatitude and λ is the east longitude, while (A,B,C,D) are complicated functions of $(\theta, \lambda, \Sigma_P, \Sigma_H, \Psi)$. This partial differential equation can be solved for Ø using reasonable boundary conditions at the pole and at the equator, at which point it is possible to calculate $\mathbf{E}$, $\mathbf{J}$, and $j_{\parallel}$ at all points in the ionosphere assuming some ionospheric conductivity distribution. Figure 1-26 shows calculated electric potential and current distributions in the high latitude ionosphere for averaged activity levels for disturbed days in May and June 1965.
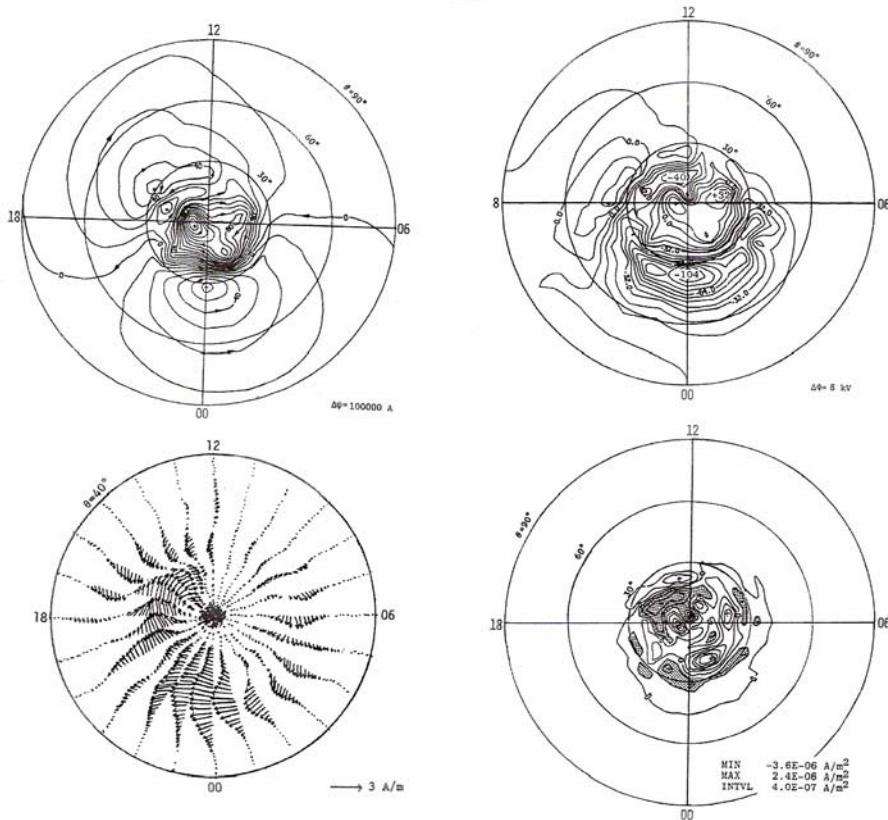
Figure 1-26: Input and Output from the KRM Routine.

Top left panel is the equivalent current system (10^5 A contour interval) obtained from ground magnetometer data for the instant 1252 UT during a substorm on January 12, 1973. The top right panel shows the pattern of electric potential (8 kV between potential contours), the bottom left panel shows the ionospheric current vectors (the length of the current vector in A/m as shown in the panel), and the bottom right panel shows the field-aligned current distribution (4×10^{-7} A/m^2 contour interval) with downward current indicated by stippled shading (after Kamide et al. [43]).

The technique described above is called the KRM method, and demands the assumption of a conductivity distribution. Since this is rarely known at any instant, it makes it difficult to apply this technique to given moments in time. Some years after the KRM technique was implemented, Richmond and Kamide [44] enhanced its capability by allowing additional information to be incorporated into the basic methodology (e.g. electric field data, conductivity data, etc.) creating the method of Assimilative Mapping of Ionospheric Electrodynamics (AMIE). The AMIE technique has been able to incorporate UV luminosity data from high altitude orbiting satellites (e.g. Polar, IMAGE) to infer more detailed information about the ionospheric conductivity distribution using the techniques described in Lummerzheim et al. [45].

1.3.4.8   Separation of Substorm Expansion Phase Effects from Directly Driven Activity

Despite the major advances in providing a quantitative description of the behaviour of key electrodynamic parameters in the magnetosphere-ionosphere system during intervals of activity, the problem still persists in being able to decouple the effects of substorm expansion phase activity from the background directly driven activity. This kind of problem is evident in examining the use of (for example) the AL index in tracking substorm activity. The AL index is a time series of points made from a superposition of the H-components of approximately 12 stations distributed as evenly as possible in longitude across all local times at average auroral zone latitudes. Each point represents the maximum of all the negative perturbations detected at the 12 stations, regardless of what local time the station that produced that maximum value is located. Negative perturbations arising from directly driven activity tend to peak near dawn, while those due to substorm expansion phase development tend to peak near midnight. Circumstances often arise in which the maximum perturbation due to directly driven activity is larger than the maximum perturbation due to directly driven activity. Then the expansion phase onset is not visible in the AL time series, which can lead to significant problems in associating disturbances measured by satellites with those observed on the ground.

An obvious solution to this problem is to examine individual magnetograms when studying any given event; however, this poses a formidable logistical problem because of the number of ground magnetograms required to optimize the likelihood of identifying the actual onset time of a typical substorm expansion phase. One way to circumvent that problem is to develop a technique that separates the effects of directly driven activity from expansion phase disturbances. This problem has been addressed by Sun et al. [46] who have developed a technique called the method of natural orthogonal components (MNOC).

The technique involves the representation of the horizontal component of the magnetic field perturbations at all available magnetometer stations over the earth's surface by an equivalent current system which is described by a function $E(t_i, r_j)$ which is a function of time t_i $(I = 1, 2, \ldots \ldots m)$ and position r_j $(j = 1, 2, \ldots n)$. The equivalent current is computed for latitudes from the pole to 50^oN at 2° intervals and for 15° intervals of longitude, so that spatially $j = 1, \ldots \ldots 480$. The equivalent current function is expressed as an expansion in terms of functions X_j^k $(k = 1, 2, \ldots \ldots h \leq n)$, viz.

$$E_{ij} = \sum_{k=1}^{h} T_i * X_k^j$$

where T_i* is a function of time and the X_j^k values are orthogonal over the observation points, viz.

$$\sum_{k=1}^{n} X_j^l \, X_j^p = 0 \quad l \neq p$$

Using a minimization procedure, and defining

$$A_{lp} = (1/m) \sum_{i=1}^{h} E_{il} \, E_{ip}$$

$$\lambda_k = (1/m) \sum_{i=1}^{m} (T_i^*)^2 \sum_{p=1}^{n} (X_p^k)^2$$

$$Xk = (X_1^k, X_2^k, \ldots X_n^k)^T$$

where (1/m) is a normalization factor, one ends up solving the equation

$$AX^k = \lambda_k X^k$$

where λ_k is a function of the order number k (=1,2,....20). It turns out λ_1 and λ_2 dominate over all the higher order values of λ_k, with λ_1 reflecting the two cell equivalent current of directly driven activity and λ_2 reflecting the one cell equivalent current of the substorm current wedge. Figure 1-27 shows a typical equivalent current system at an instant of time during substorm expansion phase activity, together with the two equivalent current patterns associated with directly driven and substorm current wedge activity as separated by the MNOC technique.
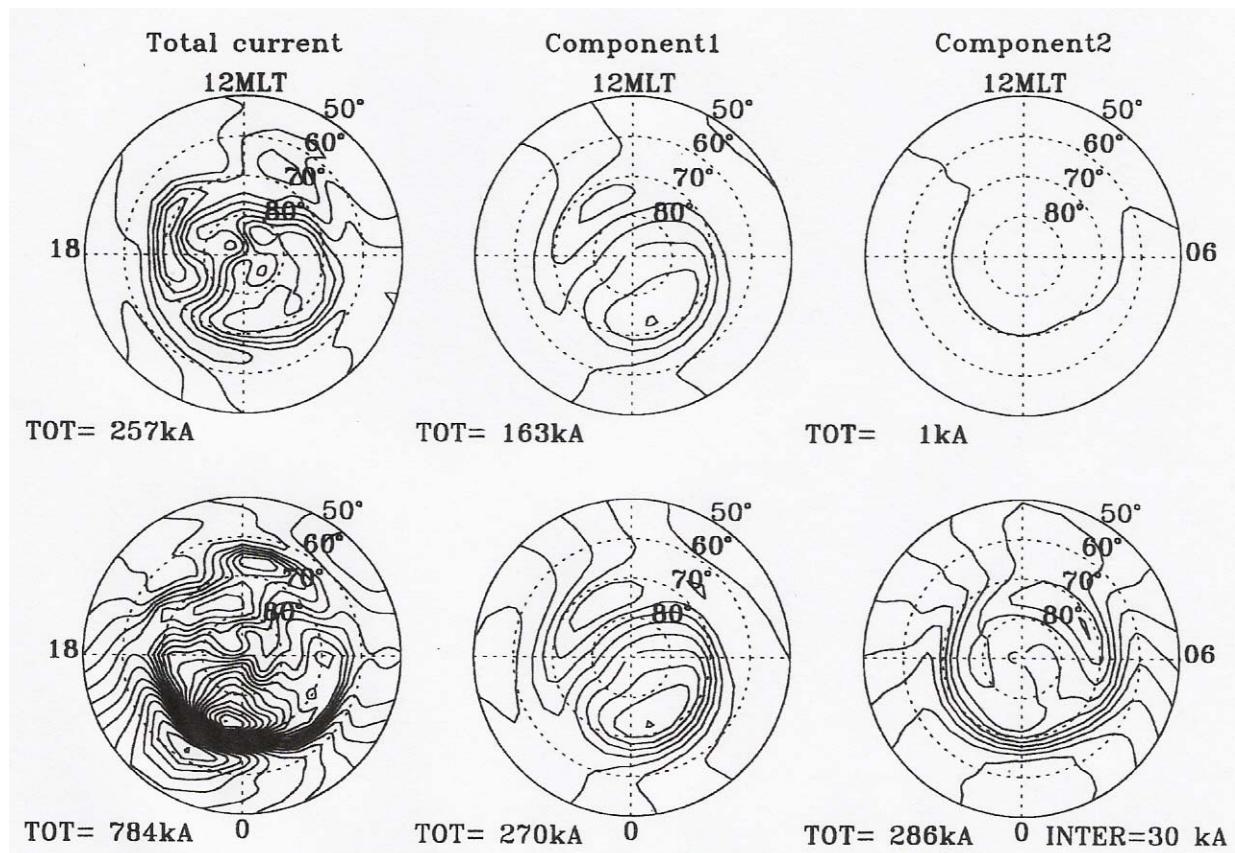
Figure 1-27: The Top Row of Polar Plots Show the Equivalent Current System for a Growth Phase Interval Followed by the Two Lowest Order Functions λ1 and λ2 Representing the Two Cell and One Cell Patterns Respectively.

The bottom row of polar plots represents the same information as the top row except for expansion phase conditions. Note, during growth phase conditions, the one cell system is virtually absent, while both one cell and two cell systems are present during expansion phase conditions (after Sun et al. [46]).

In closing, it should be said that the success of this technique depends greatly on the ground based magnetometer array coverage. If there are large gaps in the global array, this will reduce the effectiveness of the technique, since the minimization technique will tend to place the largest errors in regions where there are no data to constrain the result. However, it represents the only effective technique to date that has the potential of separating the effects of steady magnetospheric convection from effects associated with substorm expansion phase activity.

1.3.4.9 Final Comments on Substorms

In the sections above, I have outlined the concept of the substorm from the time of early observations at the beginning of the 20th century until modern times, trying to tie the terminology of the early days to the terminology employed in modern day substorm research. The original substorm definition in the context of auroral observations by Akasofu [1] has changed considerably due to the recognition that substorm activity features two characteristically different disturbances that co-exist (viz. directly driven activity and the loading-unloading behavior of the magnetotail the unloading component of which is associated with the substorm current wedge). At the same time, the recognition of the double oval structure especially evident in the evening hours has made it necessary to be specific about the nature of any auroral intensifications that are observed in the midnight to evening sector of the auroral oval. The researcher now must ask whether any observed auroral activation is a legitimate expansion phase onset, a pseudo-breakup or a poleward boundary intensification (PBI).

Another important point to recognize is that substorm research is not sufficiently advanced to permit the decoupling of effects due to directly driven activity from those due to expansion phase activity. In principle, the mathematical tools are available to attempt such a decoupling (cf. the MNOC technique of Sun et al. [1998]); however, the distribution of ground observatories whose data are needed to establish the equivalent current system for any given event at any moment in time is inadequate to provide the input to the mathematical routines that would produce believable results in detail. The same can be said of the advanced modeling techniques such as KRM (cf. Kamide et al. [43]) or AMIE (cf. Richmond and Kamide [44]) which would be extremely effective if the input data were adequate. Further progress in substorm research using these modeling tools awaits the deployment of adequate global arrays of instrumentation that will provide the input data required to produce detailed estimates of key parameters such as the cross polar cap potential drop and the strength and location of the field-aligned currents that connect the source regions of the outer magnetosphere to the regions of energy dissipation in the upper atmosphere.

Finally, it should be noted that the past decade has provided the best data set that space scientists have had to work with in attempting to develop a clear understanding of the substorm phenomenon. With continuous solar wind coverage from the Wind and ACE satellites and superb topside imaging from the Polar and IMAGE spacecraft, researchers have had the best opportunity ever to examine the details of substorm development and the response of substorm activity to the input of energy from the solar wind. This type of information, together with better ground based arrays of instrumentation, are ultimately going to be required to provide a final understanding of the role that substorms play in the solar-terrestrial interaction.

1.3.4.10 Acknowledgements

This research was supported by the Natural Sciences and Engineering Research Council of Canada.

1.3.4.11 References

[1]   Akasofu, S.-I., (1964) The development of the auroral substorm, Planet. Space Sci., 12, 273.

[2]   Birkeland, K, (1908) *The Norwegian Aurora Polaris Expedition 1902-1903*, 1, section 1, Aschhoug and Co., Christiania.

[3]   Akasofu, S.-I., (1968) *Polar and Magnetospheric Substorms*, D. Reidel, Norwell, Mass.

[4]   Dungey, J.W., (1961) Interplanetary magnetic field and the auroral zones, Phys. Rev. Lett., 6, 47.

[5]   Axford, W.I. and C.O. Hines, (1961) A unifying theory of high latitude phenomena and geomagnetic storms, Can. J. Phys., 39, 1443.

[6]   McPherron, R.L., (1970) Growth phase of magnetospheric substorms, J. Geophys. Res., 75, 5592.

[7]   Atkinson, G., (1966) A theory of polar substorms, J. Geophys. Res., 71, 5157.

[8]   Baker, D.N., et al. (1996) Neutral line model of substorms: Past results and present view, J. Geophys. Res., 101, 12,975.

[9]   Kennel, C.F., (1995) *Convection and Substorms: Paradigms of magnetospheric phenomenology*, Oxford Univ. Press, New York.

[10] Chapman, S., and Bartels, J. (1940) *Geomagnetism*, vol. 1, Clarendon, Oxford.

[11] Obayashi, T. (1967) The interaction of solar plasma with geomagnetic field, disturbed condition, in *Solar Terrestrial Physics*, p.107, ed. by J.W. King and W.S. Newman, Academic Press, London.

[12] Akasofu, S.-I. (1980) What is a magnetospheric substorm?, in *Dynamics of the Magnetosphere*, ed. by S.-I. Akasofu, p. 447, D. Reidel Publ. Co., Dordrecht, Holland.

[13] Baumjohann, W. (1983) Ionospheric and field-aligned current systems in the auroral zone: A concise review, Adv. Space Res., 2, 55.

[14] Sugiura, M., and Heppner, J.P. (1965) The Earth's magnetic field, in *Introduction to Space Science*, ed. by W.N. Hess, p. 5, Gordon and Breach, New York.

[15] Akasofu, S.-I., Chapman, S. and Meng, C.-I. (1965) The polar electrojet, J. Atmos. Terr. Phys., 30, 227.

[16] Rostoker, G. (1969) Classification of geomagnetic disturbances, J. Geophys. Res., 74, 5161.

[17] Kisabeth, J.L. and Rostoker, G. (1973) Current flow in auroral loops and surges inferred from ground-based magnetic observations, J. Geophys. Res., 78, 5573.

[18] Tighe, W.G., and Rostoker, G. (1981) Characteristics of westward traveling surges during magnetospheric substorms, J. Geophys., 50, 51.

[19] Craven, J.D., Frank, L.A. and Akasofu, S.-I. (1989) Propagation of a westward traveling surge and the development of persistent auroral features, Geophys. Res. Lett., 94, 6961.

[20] Sugiura, M. (1964) Hourly values of equatorial Dst for the IGY, *Ann. Int. Geophys Year*, 35, 9.

[21] Rostoker, G., and Fälthammar, C.-G. (1967) Relationship between changes in the interplanetary magnetic field and variations in the magnetic field at the Earth's surface, J. Geophys. Res., 72, 5853.

[22] Akasofu, S.-I., DeForest, S., and McIlwain, C.E. (1974) Auroral displays near the "foot" of the field line of the ATS-5 satellite, Planet. Space Sci., 22, 25.

[23] Burton, R.K., McPherron, R.L., and Russell, C.T. (1975) An empirical relationship between interplanetary conditions and Dst, J. Geophys. Res., 80, 4204.

[24] Gonzalez, W.D., et al. (1994) What is a geomagnetic storm?, J. Geophys. Res., 99, 5771.

[25] Kamide, Y., et al. (1998) Current understanding of magnetic storms: Storm-substorm relationship, J. Geophys. Res., 103, 17,705.

[26] Sharma, A.S., et al. (2003) The storm-substorm relationship: Current understanding and outlook, in *Disturbances in Geospace: The Storm-Substorm Relationship*, ed. by Y. Kamide et al., Geophysical Monograph 142, p. 1, American Geophysical Union, Washington, DC.

[27] Tsurutani. B.T., Zhou, X.-Y., and Gonzalez, W.D. (2003) A lack of substorm expansion phases during magnetic storms induced by magnetic clouds, in *Disturbances in Geospace: The Storm-Substorm Relationship*, ed. by Y. Kamide et al., Geophysical Monograph 142, p.23, American Geophysical Union, Washington, DC 2003.

[28] Zhou, X.-Y., et al. (2003) Ring current intensification and convection-driven negative bays: Multisatellite studies, *J. Geophys. Res.*, *108*(A11), 1407, doi: 10.1029/2003JA009881.

[29] Rostoker, G., (1996) Phenomenology and physics of magnetospheric substorms, J. Geophys. Res., 101, 12,955.

[30] Senior, C., and Blanc, M. (1984) On the control of magnetospheric convection by the spatial distribution of ionospheric conductivities, J. Geophys. Res., 89, 261.

[31] Rostoker, G. (1983) Triggering of expansive phase intensifications of magnetospheric substorms by northward turnings of the interplanetary magnetic field, J. Geophys. Res., 88, 6981.

[32] Wallis, D.D. and Budzinski, E.E. (1981) Empirical models of height integrated conductivities, J. Geophys. Res., 86, 125.

[33] Rees, M.H. (1963) Auroral ionization and excitation by incident energetic electrons, Planet. Space Sci., 11, 1209.

[34] Rees, M.H. (1969) Auroral electrons, Space Sci. Rev., 10, 413.

[35] Spiro, R.W., Reiff, P.H., and Maher, Jr., L.J. (1982) Precipitating electron energy flux and auroral zone conductances – an empirical model, J. Geophys. Res., 87, 8215.

[36] Robinson, et al. (1987) On calculating ionospheric conductances from the flux and energy of precipitating electrons, J. Geophys. Res., 98, 2565.

[37] Hardy, D.A., Gussenhoven, M.S., Raistrick, R., and McNeil, W.J. (1987) Statistical and functional representations of the pattern of auroral energy flux, number flux and conductivity, J. Geophys. Res., 92, 12,275.

[38] Ahn, B.-H., et al. (1998) An ionospheric conductance model based on magnetic disturbance data, J. Geophys. Res., 103, 14,769.

[39] Hardy, D.A., Gussenhoven, M.S., and Holeman, E. (1985) A statistical model of auroral electron precipitation, J. Geophys. Res., 90, 4229.

[40] Rostoker, G., Lui, A.T.Y., Murphree, J.S., and Anger, C.D. (1987) North-south structures in the midnight sector auroras as viewed by the VIKING imager, Geophys. Res. Lett., 14,407.

[41] Zmuda, A.J., and Armstrong, J.C. (1974) The diurnal flow pattern of field-aligned currents, J. Geophys. Res., 79, 4611.

[42] Mishin, V.M., Bazarzhapov, A.D. and Shpynev, G.P. (1980) Electric fields and currents in the earth's magnetosphere, in *Dynamics of the Magnetosphere*, ed. by S.-I. Akasofu, p. 249, D. Reidel Publ. Co., Hingham, MA.

[43] Kamide, Y., Richmond, A.D., and Matsushita, S. ([1981) Estimation of ionospheric electric fields, ionospheric currents, and field-aligned currents from ground magnetic records, J. Geophys. Res., 86, 801.

[44] Richmond, A.D., and Kamide, Y. (1988) Mapping electrodynamic features of the high-latitude ionosphere from localized observations, J. Geophys. Res., 93, 5741.

[45] Lummerzheim, D., et al. (1996) High time resolution study of the hemispheric power carried by energetic electrons into the ionosphere during the May 19/20, 1996 auroral activity, Geophys. Res. Lett., 24, 987.

[46] Sun, W., Xu, W.-Y., and Akasofu, S.-I. (1998) Mathematical separation of directly driven and unloading components in the ionospheric equivalent current during substorms, J. Geophys. Res., 103, 11695.

1.3.5   Ionospheric Troughs at F-Region Altitudes

(Alan S. Rodger, British Antarctic Survey, Cambridge, UK)

1.3.5.1   Introduction

The term ionisation trough is applied to a region of the mid- to high-latitude F layer, typically a few degrees wide in latitude wide, where the plasma concentration is low, with the plasma concentrations both poleward and equatorward of this region being higher. Troughs in the polar ionosphere are normally considered in two forms, the mid-latitude or main ionospheric trough, usually occurring just equatorward of the auroral precipitation boundary, and the high latitude trough observed within the auroral oval or polar cap. Troughs have major effects on the propagation of radio waves both because of the strong gradients in plasma concentration and also because the edges are often regions of high plasma irregularity occurrence. Troughs comprise three parts, the equatorward, and poleward edges and the trough minimum (Figure 1-28). To date there is no agreed quantitative definition of a trough making intercomparisons of results from different analyses subjective. Despite these limitations, the overall morphology of both high- and mid-latitude troughs is reasonably well known, and summarised in Rodger et al. (1992). The key features of trough morphology, and an overview of the physics of trough formation are summarised here, and some areas for further studies in order to improve the predictions of troughs characteristics are identified.
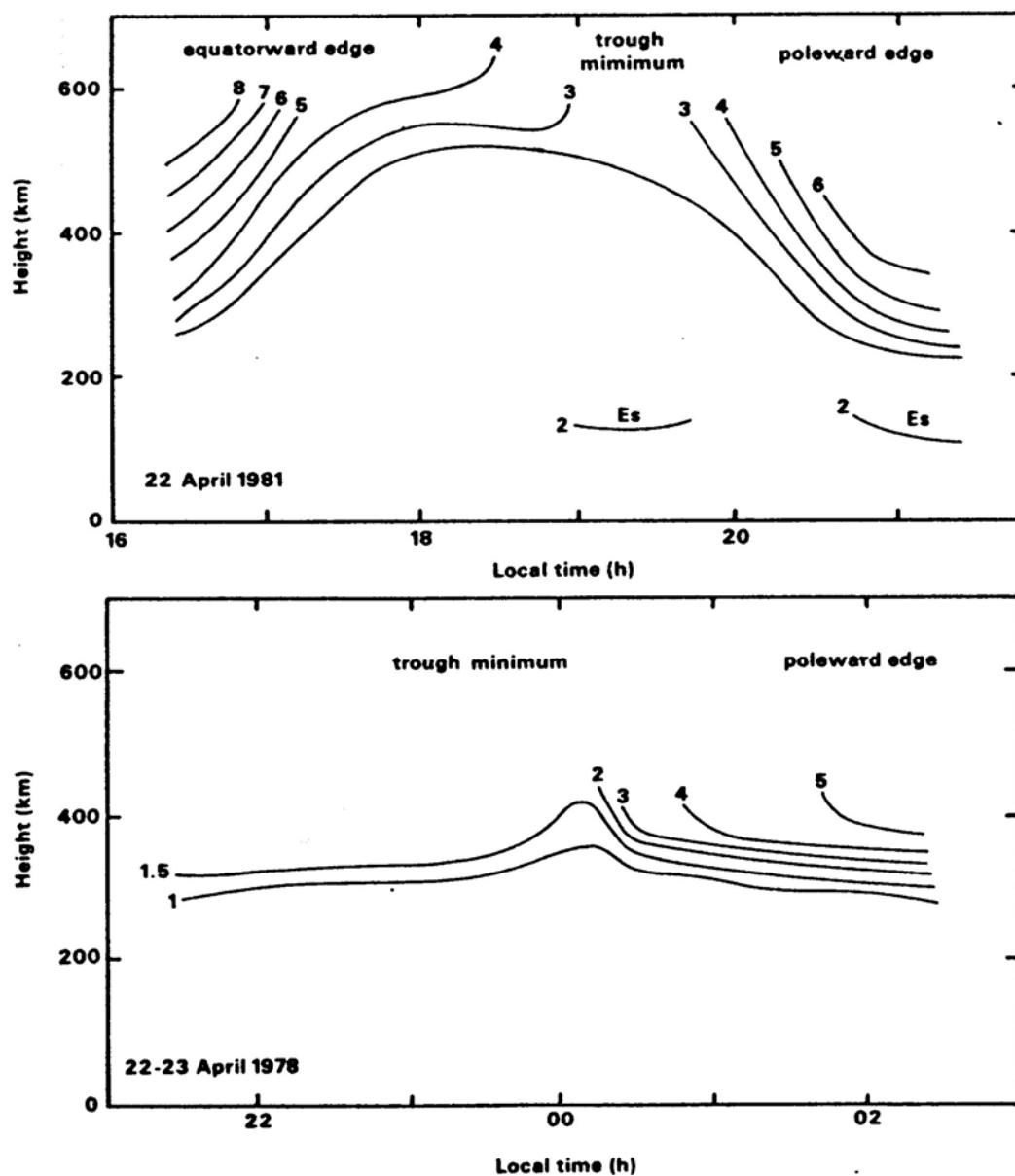
**Figure 1-28: Schematic Diagram of the Mid-Latitude Trough for Two Occasions; the upper panel
is representative of an evening sector trough with a well-formed equatorward edge,
whereas the lower panel is more typical of a morning sector trough
with an ill-defined equatorward edge. The numbers associated
with the contours indicate plasma frequency in MHz.**

1.3.5.2 Mid-Latitude Trough Morphology

The mid-latitude trough is primarily a phenomenon that occurs in darkness thus is seen mainly in winter and equinox and is less common in summer. The local time extent of the trough is small in summer, centred around midnight and extends further towards dawn and dusk with progression towards winter. In some longitudes, and under the quietest magnetic conditions, the trough can be observed at all local times around mid-winter. An excellent example of this dayside main trough is presented by Pryse et al. (2005) who show a deep trough at the equatorward edge of the auroral oval. These authors use the term 'high-latitude trough' owing to the geographic latitude, rather than the physical definition suggested above.

During steady geomagnetic conditions, the trough moves progressively to lower latitudes with later magnetic local times and is observed at lower latitudes at the same magnetic local time as geomagnetic activity increases. There are several empirical equations that describe this motion reasonably well (Dudeney et al., 1983). Most formulae use the Kp magnetic index, which has a number of limitations, an index related to the solar wind conditions might lead to improved representations. As solar activity increases the electron concentration in trough minimum and its edges also increases.

1.3.5.3   Formation Processes of the Mid-Latitude Trough – Steady State

During steady geomagnetic conditions, the trough in the evening sector forms immediately equatorward of the low latitude boundary of auroral precipitation. This is a region where the corotation and convection electric fields are oppositely directed (i.e. quasi stagnation) and hence the plasma residence time at a particular magnetic local time is greater than at lower latitudes. Provided the plasma is in darkness, then normal F-region plasma recombination processes cause the plasma to decay further compared with the slightly lower latitudes. Poleward of this region in the evening sector there are two major causes of the enhancement of plasma. First energetic particle precipitation causes additional ionisation both at E- and F-region altitudes. Sporadic E layers can be formed in addition to enhancements of the F-region concentration (see Figure 1-28). Secondly, some plasma forming the poleward edge has been transported from the dayside ionosphere, through the cusp, polar cap and nightside oval contributing to the elevated plasma concentration. Hence a trough is formed i.e. a region of low plasma concentration between two regions of higher plasma concentration.

In the morning sector, the trough minimum is the extension of the evening sector trough that is slowly convecting and/or corotating towards dawn. This convection can cause the trough to be observed slightly (~1 h) beyond the day-night terminator. The equatorward edge is formed by corotating plasma which continues to decay with time becoming progressively less distinct from the trough minimum with time (Figure 1-28 lower panel). The poleward edge is formed by plasma that has drifted across the polar cap from the dayside ionosphere with some enhancement in the plasma concentration as the flux tubes drift through the nightside oval and then to lower latitudes. The poleward edge occurs equatorward of the auroral oval and in this respect is different from the evening sector trough.

Thermosphere-ionosphere general circulation models reproduce the steady-state conditions reasonably well. This includes the variations that occur as a function of season, Universal Time (UT) (including the effects of the offset of the geographic and geographic poles), and the variations of the B_z and B_y components of the interplanetary magnetic field. The UT effects lead to particularly strong seasonal variations in trough shape, longitude extent, and gradients of the equatorward and poleward walls of the trough; these features vary as a function of solar activity.

In summary, the mid-latitude trough is a delicate balance between plasma production in the vicinity of the poleward edge of the trough, horizontal transport arising from the convection electric field around the trough minimum, and recombination processes particularly in the trough minimum and the equatorward edge. The incorporation of ionospheric outflow during the daytime and its subsequent return to the F region at night affects markedly the plasma concentration in the trough minimum, and is probably the most difficult parameters to model realistically.

1.3.5.4   Formation Processes of the Mid-Latitude Trough – Geomagnetically-Active Periods

During geomagnetically-active periods, the detailed morphology of troughs becomes much more difficult to predict although the physics is essentially identical. The fundamental problem is that the spatial and temporal variations of critical processes are not sufficiently well known to allow accurate modelling to be undertaken. The most important unknown is the spatial and temporal variation of electric fields. The precise nature of the trough morphology depends critically on the time history of the convection

pattern and hence the trajectory of plasma in the sub-auroral and high latitude ionospheres. As the lifetime of F-region plasma is many hours, especially in winter, and plasma only takes 1-2 hours to convect from the cusp into the nightside auroral oval, the electric field over the entire high latitude region is required, probably with a temporal resolution of a few minutes, to be able to predict the electron concentrations and gradients accurately.

There is a wide array of space and ground-based measurements now available in near-real time to assimilate into prediction models. Particularly valuable are the ever-increasing number of SuperDARN radars, TEC receivers, ionospheric tomography (Kersley et al., 1997) and magnetometers, but still this is insufficient information to give reliable global forecasts. Particular challenges occur where plasma velocities ($\mathbf{V}$) exceed about 500 ms^{-1} in the rest frame of the neutral wind ($\mathbf{U}$). On these occasions, enhanced plasma loss occurs (Schunk et al., 1975). It is a highly non-linear process: an increase of plasma velocity in the rest frame of the neutrals ($\mathbf{V}$-$\mathbf{U}$), from 1 to 2 kms^{-1} leads to an order of magnitude increase in the plasma loss rate. The neutral wind is often assumed to be zero but this too can lead to serious errors both in the loss rate, and in the field-aligned plasma motion. Some plasma velocity enhancements occur in narrow regions (a few km) associated with auroral arcs. Thus deep narrow troughs can be formed in minutes which are not detectable by many types of instruments, but they affect plasma concentrations and irregularity occurrence and hence radio wave propagation.

Penetration of the electric field from high to mid and low latitudes is one of the areas of most active research at present and a topic that has great relevance for the formation and morphology of the mid-latitude trough. A wide variety of terminology for the phenomenon has been used over the last three decades, including polarisation jet (Galperin et al. 1973), sub-auroral ion drift (SAID) (Spiro et al., 1979), substorm-associated radar auroral surges (SARAHs) (Freeman et al., 1992), and sub-auroral polarisation streams (SAPS) (Foster and Burke, 2002). The plasma jets not only are regions of enhanced loss, but they also transport low concentration, night-side plasma into the dayside ionosphere. As geomagnetic increases and the auroral oval and convection patterns expand, high plasma concentration mid-latitude plasma is entrained into the high-latitude convection pattern and lead to storm enhanced densities (SEDs), particularly in the afternoon sector (Foster et al., 2002). GPS measurements of total electron content have helped very significantly in determining the spatial and temporal variations of plasma concentrations during storm time but accurate prediction of the penetration fields is still some way off.

Energetic particle precipitation is particularly important in forming the poleward edge of the mid-latitude trough. As with electric fields, the general morphology is well known but the detail of where, when and the spectrum of the precipitating particle flux is less predictable. As a simple example, the first substorm normally occurs within about 1 hour after a southward turning of the interplanetary magnetic field but the longitude of initiation and the extent of the substorm cannot yet be predicted, so the pattern energy deposition from the magnetosphere into the ionosphere cannot be predicted.

Energetic particle precipitation also leads to increased conductivity, which in turn leads to enhanced Joule heating in the presence of an electric field. This is important for trough formation in two ways. First it can increase or decrease the neutral wind locally causing the altitude of the plasma to change, and hence affect the lifetime of plasma. Also regions of high Joule heating is where there is rapid ion outflow which contributes to plasma depletion at F-region altitudes i.e. a trough. These factors contribute significantly to the difficulty in predicting accurately trough characteristics.

1.3.5.5   High Latitude Trough Morphology

The morphology of high latitude troughs is less well defined mainly because of the less comprehensive sets of ground-based observations in very high latitudes due to the logistic difficulties of establishing instrumentation near both poles. Satellite data have played a key role in trough identification. Figure 1-29 gives an example of a trans-polar crossing showing trough structures both within the polar cap and the

auroral oval. This figure also illustrates the problem of having no agreed definition of a trough, as many trough-like features can be identified. The physical processes for high-latitude trough formation are essentially the same as those of the mid-latitude trough, i.e. a very delicate balance between various plasma transport processes, production and loss of plasma and again the combination of them is unique on each occasion.
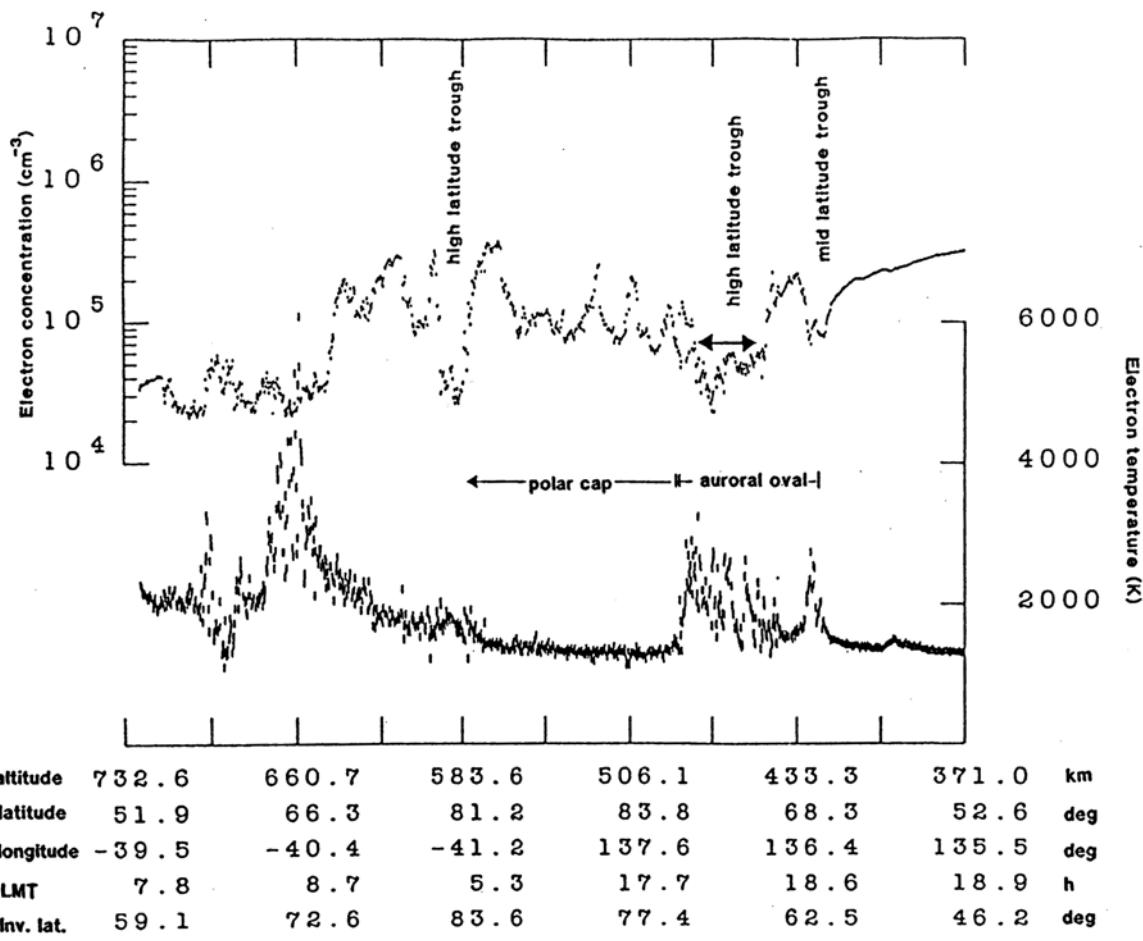
Figure 1-29: Latitude Profile of Electron Concentration (left-hand scale) and Electron Temperature (right-hand scale) from Dynamics Explorer 2 Spacecraft, 22 November 1981 Showing the Relationship of High-Latitude Troughs with Respect to the Mid-Latitude Trough, Auroral Oval and Polar Cap.

The most familiar high-latitude troughs are those that form deep in the polar cap when plasma convection trajectories do not experience sunlight for very extended intervals (many hours) and plasma concentration in the F region decays with time – these are called polar holes.

High-latitude troughs have also been found to be associated with the boundary between a lobe cell and a viscous convection cell during periods of IMF northward conditions (Pryse et al., 2006). This is a region where there is a large horizontal velocity shear, and hence field-aligned currents and intense energetic particle precipitation. Adjacent convecting flux tubes have a very time history, which can result in trough-like features at F-region altitudes. These effects can be sufficiently large that troughs can be observed even when the in the high-latitude ionosphere is in total sunlight in summer (see Figure 1-30 for example).
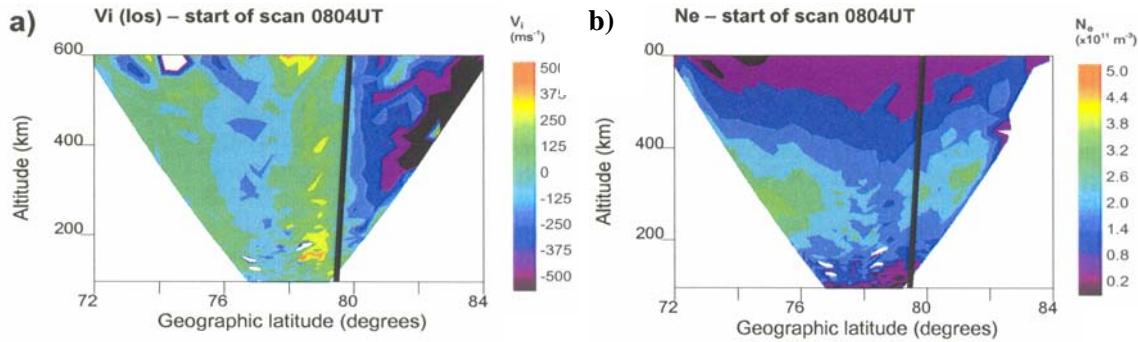
**Figure 1-30: The Line-of-Sight Velocities Determined from the EISCAT Svalbard Radar at 0804
UT on 24 August 2000 with the Corresponding Electron Concentrations. The black
line marks the centre of a significant velocity shear with a high-latitude
trough lying close to the shear (after Pryse et al., 2005).**

Considerable structuring of F-region plasma occurs in the vicinity of the polar cusp. High-concentration plasma enters the polar cap on the dayside, but can become highly structured by the effects of variable, and large magnitude (>1 km s^{-1}) electric fields, resulting in the formation of polar patches (Rodger et al., 1994a). These structures persist right across the polar cap and enter the nightside oval (Rodger et al., 1994b).

A further way of forming troughs in the polar cap is the effects of high plasma velocities (500 ms^{-1}) associated with trans-polar arcs that cause rapid recombination and hence reduce the plasma concentration at F-region altitudes (Schunk et al., 1975).

The relative importance of these various mechanisms, or the precise geophysical conditions under which they occur has yet to be fully established.

1.3.5.6   Focus for the Future

Scientific progress is often made by the inter-linking of four approaches. These are observations (the crucial link to reality), data assimilation techniques, sophisticated modelling developments reflecting an ever-increasing understanding of the underlying science, and new scientific theories. Improved trough predictions require some contribution from each of these elements.

There are ever-increasing quantity of high-quality ionospheric, magnetospheric and solar wind observations available in near-real time that can be assimilated into mature general circulation models. However observations of key parameters, such as electric fields, are not likely to be available at a sufficiently small scale to describe very detailed plasma motions. An alternative approach would to parameterise the electric field using the knowledge that the amplitude of electric field fluctuations in the polar cap can be described by a power law relationship that is very similar to that of the Poynting flux at 1 AU in the solar wind (Freeman et al., 2000). Variations of the electric field within the auroral oval also can be described by a power law but with a different exponent (Abel, private communication). However there are significant challenges to know how to relate the variations in electric field to those of ionospheric conductivity, recognising, for example, that the peak in the electric field associated with an auroral arc is not collocated with the peak conductivity. This makes the accurate calculation of Joule heating particularly challenging.

The second area that is in most need of scientific progress is a quantitative understanding of a few key processes, the most important of which are probably energy deposition as a function of space and time in a substorm, and the geophysical factors that control the penetration of the electric field from high to lower latitudes and flow into and from the topside ionosphere. Substorms have been studied for half a century

but the forthcoming THEMIS spacecraft mission with very comprehensive, complementary ground-based measurements offer some new possibilities to understand the instability processes that initiate substorms in the tail of the magnetosphere and hence predict energy deposition. The spatial/temporal variations of the region 2 and ring current regions must hold the key to the penetration of the electric field to lower latitudes. Modelling the ionospheric outflows needs further work to understand the acceleration processes at low altitude, and more realistic treatment of the upper boundary condition of thermosphere-ionosphere general circulation models is also necessary.

1.3.5.7   References

Dudeney, J. R., A. S. Rodger and M. J. Jarvis, Radio studies of the main F-region trough in Antarctica. Radio Science, 18, 927-936, 1983.

Foster, J. C., P. J. Erickson, A. J. Coster, J. Goldstein and F. J. Rich, Geophys. Res. Lett., 29, 10.1029/2002GL015067, 2002.

Foster J. C. and W. J. Burke, SAPS: a new categorization of sub-auroral electric fields, EOS. Transactions, American Geophysical Union 83, 393-394, 2002.

Freeman, M. P., D. J. Southwood, M. Lester, T. K. Yeoman and G. D. Reeves, Substorm-associated radar auroral surges, J. Geophys. Res. 97, 12173-12185, 1992.

Freeman M., N. W. Watkins and D. J. Riley Power law distributions of burst duration and interburst interval in the solar wind: Turbulence or dissipative self-organized criticality? Physical Review E 62 (6): 8794-8797 Part B, December 2000.

Galperin, Y. I., V. L. Khalipov and A. G. Zosimova, Direct measurement of ion drift velocity in the supper ionosphere during a magnetic storm, 2. Results of measurements during the November 3, 1967 magnetic storm, (In Russian), Cosmicheskie Issledovanija, 11, 273, 1973.

Kersley, L., S. E. Pryse, I. K. Walker, J. A. T. Heaton, C. N. Mitchell, M. J. Williams and C. A. Wilson, Imaging of electron density troughs by tomographic techniques, Radio. Sci. 32, 1607-1621, 1997.

Pryse S. E., K. L. Dewis, R. L. Balthazor, H. R. Middleton and M. H. Denton, The dayside high –latitude trough under quiet geomagnetic conditions: radio tomography and the CTIP model, Ann. Geophys. 23, 1199-1206, 2005.

Pryse S. E., R. W. Sims, J. Moen and K. Osavik, Ionspheric signatures of low-latitude the boundary layer under conditions of northward IMF and small clock angle, Ann. Geophys. 24, 2169-2178, 2006.

Rodger, A. S., R. J. Moffet, and S. Quegan, The role of ion drift in the formation of ionisation troughs in the mid- and high-latitude ionosphere – a review, J. Atmos. Terr. Phys., 54, 1-30, 1992.

Rodger, A. S., M. Pinnock, J. R. Dudeney, K. B. Baker, R. A. Greenwald, A new mechanism for polar patch formation. J. Geophys. Res., 99, 6425-6436, 1994a.

Rodger, A. S., M. Pinnock, J. R. Dudeney, J. Watermann, O. de la Beaujardiere and K. B. Baker, Simultaneous two-hemisphere observations of the presence of polar patches in the night-side ionosphere, *Annales Geophysicae*, 12, 642-648, 1994.

Schunk, R. W., P. M. Banks and W. J. Raitt, Effects of electric fields and other processes upon the night-time high latitude F layer, J Geophys, Res., 80, 3121-3130, 1975.

Spiro, R. W., R. A. Heelis and W. B. Hanson, Rapid sub-auroral ion drift events observed by Atmospheric Explorer C, Geophys. Res Lett. 6, 660-663, 1979.

1.3.6 Small Scale Irregularities at High Latitudes

(J.-P. St.-Maurice, Institute of Space and Atmospheric Studies, University of Saskatchewan; A.M. Hamza, Department of Physics, University of New Brunswick)

The text for this section was been provided as a separate PDF document – *click here to view Section 1.3.6.*

1.3.7 Characterizing the Polar-Cap Ionosphere

(Herbert C. Carlson – CASS, Utah State University, 4405 Old Main Hill, Logan UT, 84322-4405 and AFOSR Chief Scientist, 875 N Randolph St, Arlington VA, 22203-1954)

1.3.7.1 Introduction and Motivation

Here we address polar-cap phenomena, notably polar-cap patches and sun-aligned arcs, as they affect the ionosphere. The intent is to summarize current knowledge without the extensiveness of a full review. Several full reviews of knowledge of polar-cap patches and polar-cap sun-aligned arcs have appeared in the last decade. For patches and blobs the reader is referred to Crowley (1996). For polar-cap patches and arcs, Radio Science special section on CEDAR HLPS/STEP GAPS, 31, 573-678, (1996); Basu and Valladares, (1999). For sun-aligned arcs, Smith and Lockwood, (1996); Zhu et al., (1997); Sandholt et al., (2002). Also, Space Weather issues in general are discussed by Schunk and Sojka, (1996).

Because understanding of the polar-cap ionosphere is of practical importance to communication, GPS navigation, and satellite radar imagery, description of its phenomena and physics here includes reference to how research advances improve our ability to utilize or circumvent its effects on radio frequency propagation.

It was in fact the launch of a polar-orbiting satellite, experiencing communication outages, which led to the widespread recognition that the polar-cap ionosphere could be of any consequence to radio-frequency signals from satellites. Such outages had been predicted five years earlier by a small group of researchers, based on theoretical extrapolation of sunspot minimum data. However, without actual sunspot maximum data, the prediction could not be definitively proven; that is, not until the other half of the 11-year solar cycle had passed. A robust research program based on an experiment-theory-model has since led to actionable predictions, and operations in the polar cap have benefited. Improved definition of morphology, correlations and patterns, and understanding leading to physical models, now-casts, and forecasts, have aided design, and enabled mitigation and circumvention of many operational issues.

To understand the polar ionosphere, it must be viewed as alternating between two fundamentally different states. It changes its basic character in response to reversals of the sign of the north-south component (Bz) of the interplanetary magnetic field (IMF). The other components (By, Bx) of the IMF have important but lower-order impact on its properties. Changes can occur over scales of minutes or less through the IMF control of phenomena within the polar cap. These rapid changes are perhaps what most distinguish the polar-cap ionosphere from the rest of the globe, with which it shares other strong dependencies on the solar cycle, and on seasonal and diurnal variations. Distinction blurs, though, between the latter two when deep in the polar cap, where there are months of continuous darkness or daylight at midwinter and midsummer.

To quantify the degree of degradation of radio frequency signals from satellites, see Figure 1-31. It shows eight years of polar-cap scintillation data from Thule, Greenland, monitoring 244 MH signals from a

USAF polar beacon satellite, at high elevation angles which do not accentuate ionospheric scintillation effects. Fifteen-minute fade depths exceed 10 db over 50% of the time for each of the five winters sampled during the sunspot maximum years 1979-1982, but are present less than 5% of the time for sunspot minimum years 1985-1986. The fades reach 20 db at least 10% of the time in mid-winter for these same five solar max years.
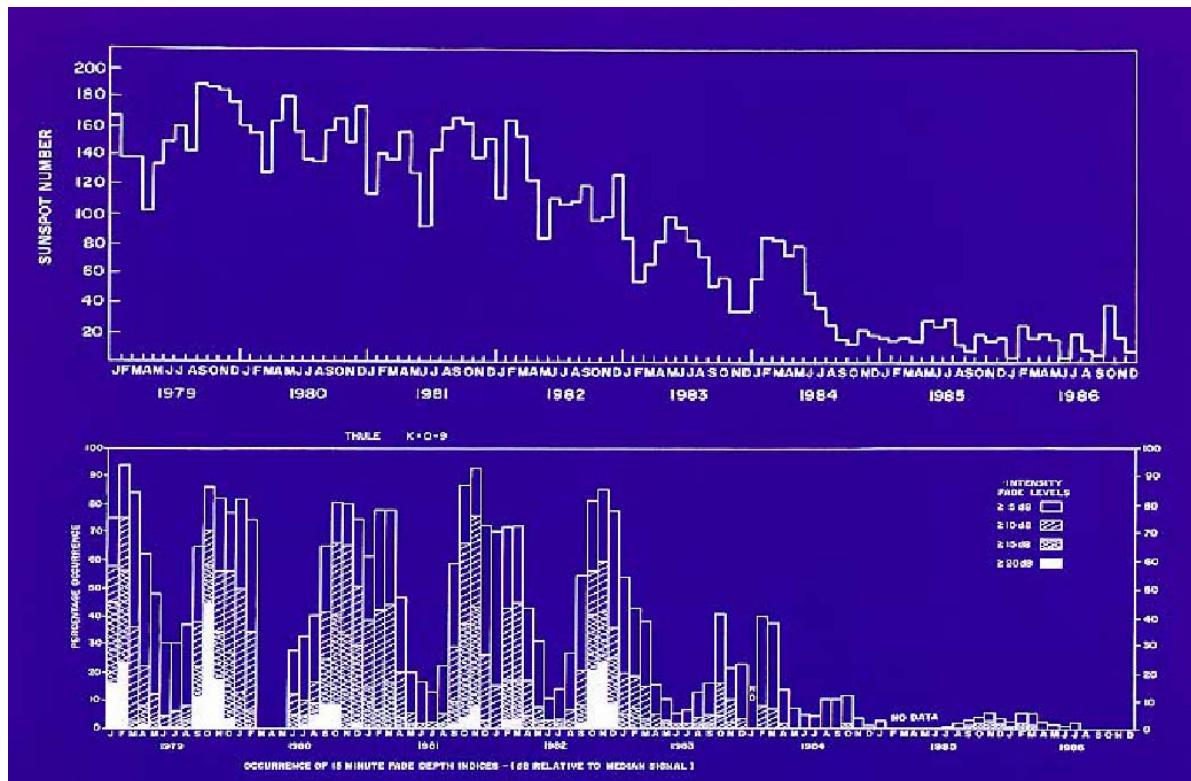
Figure 1-31: Polar-Cap Scintillation Variation with Solar Cycle, for 244 MHz Signals from a High-Elevation USAF Polar Beacon Satellite, Received at Thule, Greenland During 1979-1986. Sunspot number is in the upper plot, percent occurrence of 15-minute fade depth indices are in the lower plot for 5, 10, 15, and 20 db fade relative to median signal. (Basu et al., 1988)

1.3.7.2   The Two States of the Polar Cap

The polar ionosphere and upper atmosphere alternate between two states, depending on whether the IMF is southward or northward. This alternation is due to the very different degree of coupling between the solar wind and the magnetosphere, for opposite signs of the Z component of the IMF B_z. For southward IMF, called negative B_z, the coupling is very strong, and such strong coupling can have severe effects on radio signals. Ground-based optical and radio measurements in the polar cap respond in a way so reliably and so strongly coupled that their measurement can show which of these two states prevails. This correlation was established through the use of satellites measuring the full IMF vector (and solar wind speed, density and temperature). Satellites "parked" at the location where earth's and sun's gravitational fields balance are upstream of earth in the solar wind, and can therefore give us early warning of IMF changes about an hour before corresponding conditions switch at earth.

Below we summarize the salient optical and radio signatures that characterize the southward and northward states.

1.3.7.2.1   Southward IMF State

The state for southward IMF is characterized in midwinter by ~100-1000 km islands, or patches, of enhanced F-region plasma, surrounded by lower-density plasma. As these patches pass overhead, the overhead ionosphere can switch sharply between peak plasma densities $~10^{11}$ m^{-3} to $~10^{12}$ m^{-3} ($~10^5$ cm^{-3} to $~10^6$ cm^{-3}) (Weber et al., 1984), values which correspond to an ionospheric f_oF_2 of 3 to 9 MHz in the HF band. This excursion is literally from night to day (from typical electron densities found near midnight to near midday at mid-latitudes). The term "patches" derives from their patchy appearance in 630.0 nm imagers, which discovered them (Weber and Buchau, 1981). By convention (Crowley, 1996) high electron density islands, double or more their surrounding densities, are termed patches. Patches move antisunward across the polar cap, at ~1 km/s from near noon toward midnight. A patch formed over Svalbard, Norway, passes over Alaska about an hour or two later. Its transpolar trajectory is taken to be consistent with a normal two-cell convection pattern (Heelis, 1984) defined by averaged flow patterns of ionospheric plasma in the polar cap for southward IMF.

Figure 1-32 shows a representative patch (Weber et al., 1984). The upper sequence of frames shows 630.0 nm intensity images overhead. The field of view is about 155° of an overhead hemisphere in this fish-eye view of the All-Sky Imaging Photometer (ASIP). The bright form moving from noon (bottom of first frame) toward midnight (top of fifth frame) is the signature of the high-density plasma patch. That it is associated with high-density ionospheric plasma is established by its coincidence with approaching, then passing overhead, then receding, Doppler-shifted HF echoes as shown in the lower frames of the figure. Approaching plasma, as indicated by upward Doppler shift, is in the lower set of ionogram panels; receding plasma is in the upper set of ionogram frames. The enhanced plasma is of an ionospheric f_oF_2 of ~9 MHz, the background ~3 MHz. The 630.0 nm emission is due to the chemically-stored energy released by dissociative recombination of the ionospheric plasma, which first excites the lowest excitation state of atomic oxygen (OI)^{1}D upon recombination, then within about a minute emits a 630.0 photon as the excited state relaxes back to its ground state. The 9 MHz ionosphere has ~ten times the number of recombining electron-ion pairs per unit time, hence about ten times the atomic oxygen (OI)^{1}D excitation rate as the 3 MHz, and thus is ~ten times as bright in the 630.0 nm emission. Hence the bright patch. It is not aurora, it is just standard atomic oxygen recombination red-line airglow from nighttime F-region ionospheric recombination. For the nominal 250 km altitude of this airglow, the ASIP field of view is ~1000 km.

DRIFT OF POLAR CAP IONIZATION PATCH

THULE, 22 JANUARY 1982

6300 Å ALLSKY PHOTOMETER IMAGES

SPLIT DOPPLER IONOGRAMS
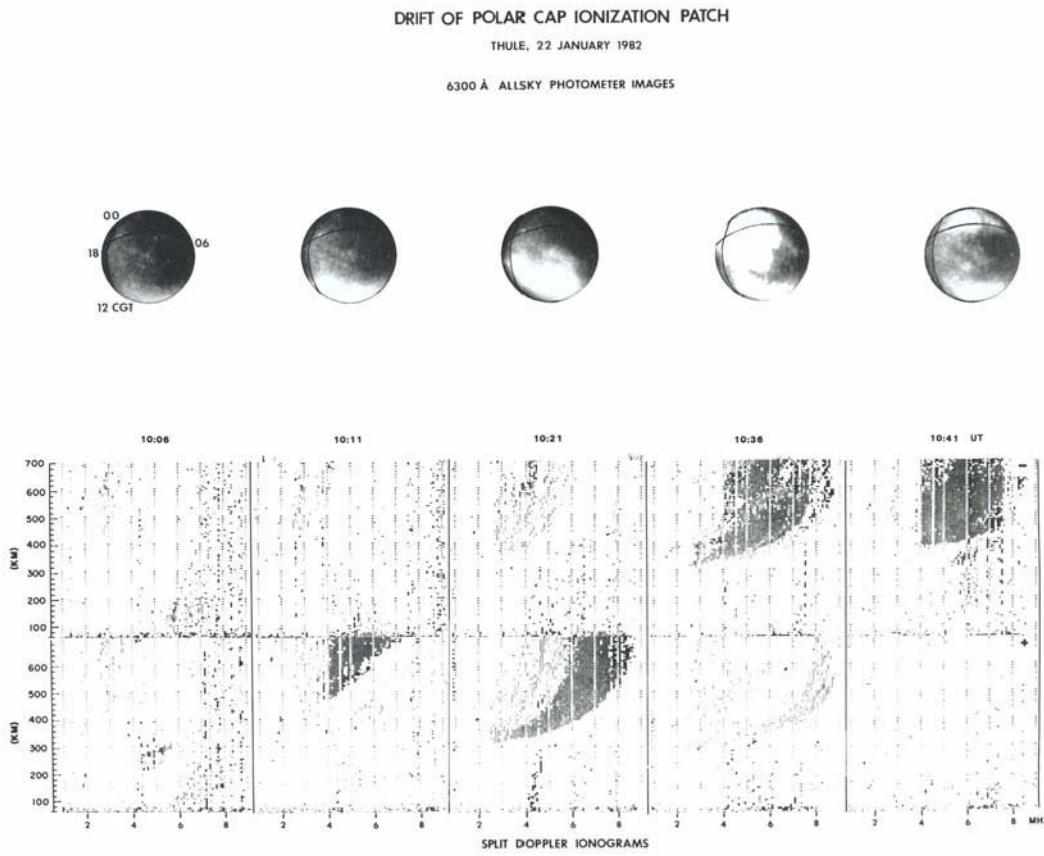
Figure 1-32: The Overhead Passage of the 630.0 nm Bright Patch Seen on the ASIP Coincides with the Overhead Passage of the High-Density Plasma Enhancement (from 3 to 9 MHz f_oF_2) Seen on a Digisonde Sensitive to Doppler Intensity and Range Discrimination (Weber et al., 1984).

The plasma within these enhanced plasma patches is highly structured over scale sizes spanning more than 0.l to tens of km, producing severe scintillation (Basu et al., 1988). These electron-density irregularities are produced by a plasma instability, the gradient drift instability, with the horizontal electrical force acting across the density gradients. The patches are less obvious in summer when the patches, still present, are of lower contrast against the higher background sunlit F-region polar plasma (Buchau et al., 1985). They are also far less structured, because the conducting E-region plasma largely shorts out the instability process.

It is thought that these patches exit the polar cap near midnight, to become the "blobs" (Vickrey et al., 1988) often seen on winter nights, in the post-sunset subauroral return-flow region. These cigar-shaped high plasma density features are also highly structured in plasma-density irregularities. Thus they represent a source of transient rf disruption, just equatorward of the polar-cap boundary.

1.3.7.2.2   *Northward IMF State*

The other state of the polar-cap ionosphere is characterized in optical emission by sun-aligned arcs. Sample observations, to contrast arcs with patches, are illustrated in the top pair of panels of Figure 1-33. The northward IMF panel shows a sun-aligned arc emerging from the dawn edge of the auroral oval and moving duskward toward the central polar cap. The motion is tracked at equal five-minute intervals per frame over a half-hour period. The arc moved steadily, persisting nearly an hour before finally starting to

fade. The bright edge at the left of the frame is the poleward edge of the polar cap (highly contracted for these northward IMF conditions) near dawn. The field of view is as in Figure 1-32. Observations from selected pairs of stations within the nighttime polar cap show that single discrete sun-aligned emissions can extend very large distances, > 2000 km for the nominal 250 km of typical F-region 630.0 nm emission, in the noon-midnight direction.
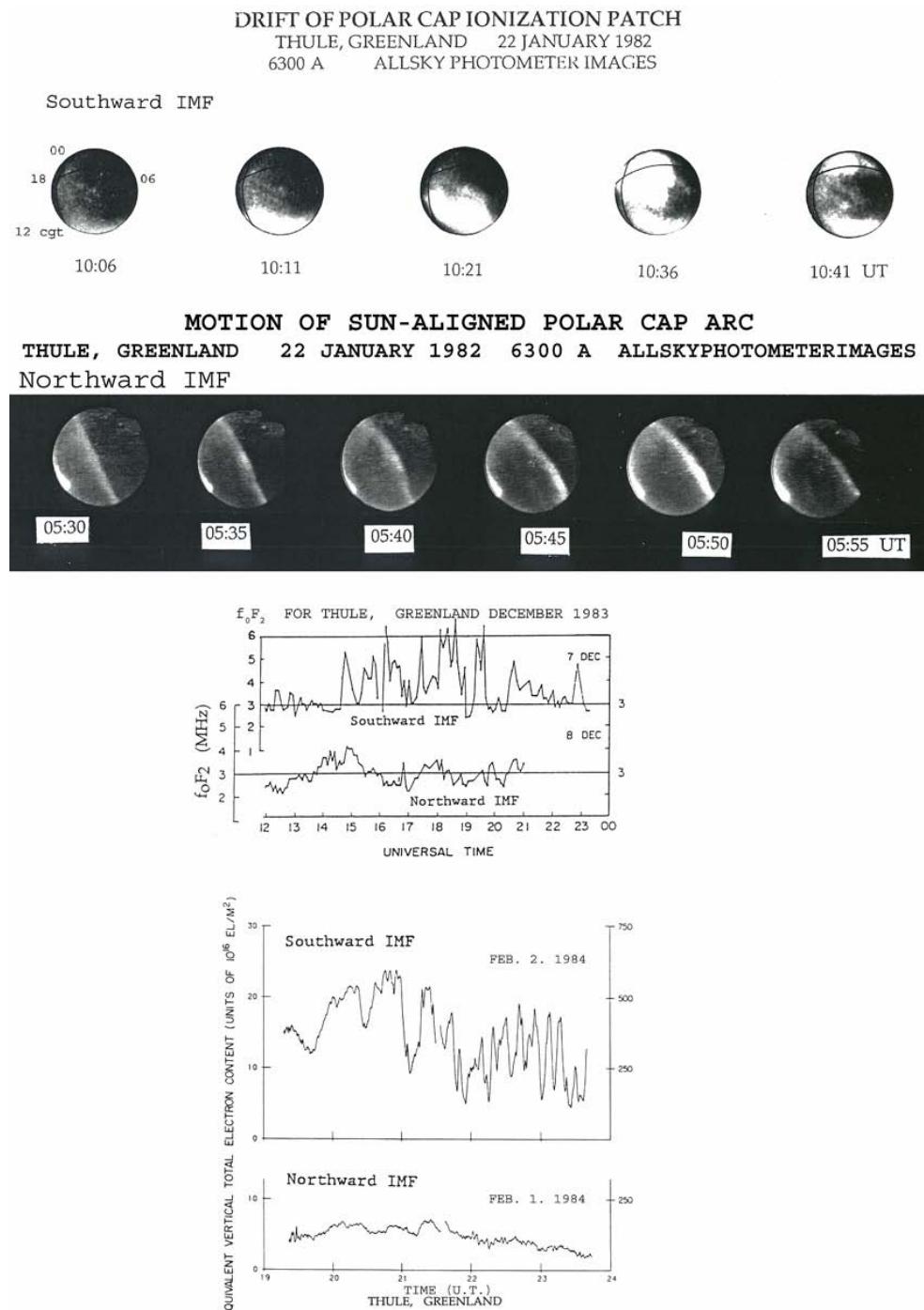
Figure 1-33: This Figure Contrasts the Signatures of the Southward vs. Northward IMF Conditions of the Polar-Cap Ionosphere – in, respectively, the upper vs. lower portions of each panel pair from: ASIP images, a digital ionosonde, and a TEC polarimeter.

The data are all overhead Thule, Greenland, in the central polar cap, during local winter. The ASIP shows an antisunward drifting patch in top panel vs. dawn-dusk drifting sun-aligned arc. For IMF south the digisonde shows sporadic strong enhancements of f_oF_2 (up to ~9MHz) above the low background nighttime level of ~3 MHz, and the TEC data shows a series of TEC enhancements about an average near 5 TEC units, both signatures of patches. For IMF north, the levels are relatively flat near the baseline f_oF_2 and TEC values on which patch enhancements sit for southward IMF. Strongly enhanced scintillation due to plasma irregularities of 0.1 to tens of km scale size coincide in time and space with overhead patches seen by any of the three diagnostics for southward IMF, and with sun-aligned arcs for sunspot maximum northward IMF conditions.

Complementary measurements by other ground-based and satellite sensors have allowed us to determine that the sun-aligned arcs are excited by sun-aligned sheets of precipitating electrons, which mark shear lines of transpolar ionospheric plasma flow. The fact that the arcs are the optical signature of electron-impact excitation of the upper atmosphere by suprathermal electron current-carriers, means that the sun-aligned arcs are true aurora. Their low intensity (100s of R) is due to weak fluxes of soft, low-energy electrons, order hundreds of eV. The state of the polar-cap ionosphere for positive B_z is characterized by considerable structure along the dawn/dusk direction in its plasma velocities, and weak but stable sun-aligned arcs seen in optical emissions.

Sun-aligned arcs were initially thought to be present only a few percent of the time (Ismail et al., 1977; Gussenhoven, 1982; Frank et al., 1986) based on observations with a sensitivity threshold of order I kR. Now we know they are prevalent for northward IMF, present in that half of the time when B_z is positive. Studies have led them to be known now as markers of highly anisotropic transpolar velocity gradients (Carlson et al., 1988). Antisunward plasma flow, decreasing (or even reversing) along a dawn-to-dusk direction, requires a vertical upward current to maintain divergence-free current across the arc. The arc emissions are impact-excited by the downgoing current-carrying electrons. Near the central part of the polar cap, the arcs drift duskward (dawnward) for positive (negative) IMF B_y at a non-uniform speed ~0.1 - 0.2 km/s) (Weber et al., 1984; Frank et al., 1986; Valladares et al., 1994). The F-region plasma above these arcs is also structured, in this case driven by a different class of plasma instability, a shear-driven instability (Basu et al., 1998). Scintillation is strong over these arcs only during sunspot maximum. During sunspot minimum, the F-region densities are much lower and the scintillation much weaker. These arcs are presumed to connect to the dayside auroral oval and are known to connect to the midnight auroral oval, although most observations are confined to the central-polar regions, with very little data available for the area of the dayside auroral region. Above the optical signature of the arcs, current sheets must then extend along the earth-sun direction, over the quasi-transpolar length of the arc, reaching upwards into the magnetosphere to as far as the electron current source.

1.3.7.2.3   *Framework for Character and Morphology*

A framework for considering the polar-cap ionosphere in terms of northward vs. southward IMF is summarized in Figure 1-33 and Figure 1-34. Figure 1-33 shows what one sees from a ground-based or airborne polar-cap location on earth, i.e., the near-earth frame of reference. Figure 1-34 shows how the field of view of Figure 1-33 fits into the context of temporal dynamics across the full polar cap, as viewed from above the polar cap in the earth-sun frame of reference.
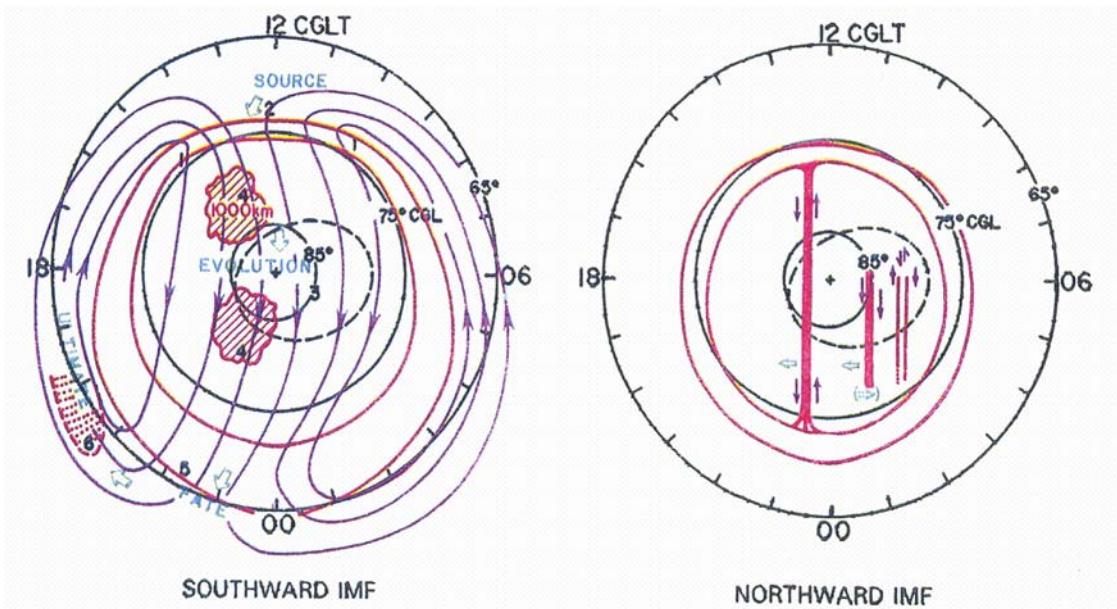
**Figure 1-34: View in Earth-Sun Frame of Reference of Conditions Characterizing the State
of the Polar-Cap Ionosphere for Southward and Northward IMF and Positive B_y.**

For negative B_y the patterns are mirror-imaged about a noon-midnight axis of symmetry. A nominal auroral oval location is shown for each condition, on CGLT coordinates from 65o magnetic latitude, looking down on the north magnetic pole. Convection is shown by closed flow lines with arrows overlaid (southward IMF) or thin adjoining arrows (northward IMF). Cross-hatched areas show recombination 630.0 nm-enhancements from high plasma density patches (southward IMF); sun-aligned bars show electron impact 630.0 nm-enhancements (northward IMF). Open arrows show the direction of motion of these polar 630.0 nm features. The dashed-line ellipse near the central polar cap is the field of view of an ASIP at the latitude of Thule, for 250 km emissions. Note that for this CGLT coordinate system, fixed magnetic latitudes are circles, so circular fields of view appear as ellipses. (Carlson, 1994)

All these descriptions are for dark northern hemisphere polar-cap winter conditions. The same physical phenomena must occur in the summer sunlit pole, but the degree of change in ionospheric properties and signatures will be largely masked and/or subdued by the ionospheric plasma locally produced by the sunlight (EUV for the F-region ionosphere). Plasma transported from large distances dominate the character of the winter conditions, whereas locally-produced plasma dominate the summer conditions.

The polar ionosphere for southward, vs. northward IMF, is characterized by quite different properties of the overhead sky seen by: an ASIP, a digital ionosonde (digisonde), or a total electron content (TEC) polarimeter receiver of signals from a satellite radio beacon. Thus, such ground-based measurements can serve as surrogate indicators of the north-south sense of the IMF.

1.3.7.3   Outline of Effects on rf Signals

The primary concerns at HF and VHF are clutter, scatter, and multi-path. Whereas, at and above VHF through GHz, they are amplitude and phase scintillation, leading to signal drop out and loss of phase lock.

Patches (and sun-aligned arcs at sunspot maximum) affect radio frequency signals at all wavelengths, so frequency diversity is of little value. The higher the frequency, the weaker the scintillation. Both scatter and scintillation stem from gradients in refractive index at a wide range of scale sizes, the relevant refractive index gradients here being due to electron-density irregularities over scales of meters (and less) to ~10s km. There exist good databases on irregularity scale size spectra, and scintillation index vs.

frequency, for a wide range of locations, time of day and season, and phase of solar cycle, e.g., Basu et al. (1988). Fades in amplitude and phase rotation rates are strong enough to lose lock on GPS signals.

Scintillation produced by both patches and arcs, which have relatively steep spatial boundaries, tends to switch on/off as a patch or arc moves across the line of sight from a receiver to a satellite rf beacon. Spatial diversity is the best mitigation (other than very large fade margins), and progress on patch characterization, location tracking, now-cast, and prediction can considerably improve system reliability. The concept, in a real-time patch-tracking scenario, is to anticipate and switch to another satellite with a free line-of-sight. Switching from one satellite to another, as different patches drift through different lines-of-sight, can significantly improve continuity of signal. Likewise for arcs. Even for "blind" operations, knowledge of the statistics on typical scale sizes and velocities of patch (and arc) envelopes, frequency of occurrence, and spatial separations, can guide message/backup strategies. Readily available real time IMF monitoring can refine this.

At HF, even at times of season when clutter seems impossible to deal with, novel ways to circumvent the issue can emerge from an understanding of patch/arc morphology and size/location of the auroral oval with time. If uninterrupted HF communication is a priority, a third station, adequately far equatorward of the auroral oval, can circumvent polar patch HF scatter problems, even while direct point-to-point communication may be impossible. The solution could be propagation via two legs of a triangle. Polar transmission to a transceiver/repeater, farther equatorward than half the distance to the far edge of the auroral oval, could give reflection off a smooth ionosphere to a third station which retransmits to the intended polar receiver. Patches and arcs can also add an HF communication path, provided the patch or arc location is adequately known or forecast. Patches and arcs move in well-defined directions with nominal steady velocities. Sun-aligned arcs can provide a sun-aligned slowly-moving reflecting wall, with persistence up to order an hour.

1.3.7.4   Patches: Theory and Modeling Context

To model patch behavior, one needs to be concerned with the source reservoir of plasma for patches; a mechanism for patch creation; the transpolar trajectory of patch path after entry into polar cap; the structuring of patch from creation to exit; the exit process for the patch; and its return sunward flow, with the possibility of recycling.

First though, geomagnetic vs. geographic coordinates are essential to patch morphology. The frequency of patch occurrence in the polar cap has been studied from ground stations measuring F-region peak electron density and total electron content, and by satellites passing through the topside F region. Rodger et al. (1994) give a good summary of the characteristics of patches. Many ground-based statistics have been published (Buchau and Reinish, 1991; McEwen and Harris, 1996; Dandekar and Bullett, 1999). Figure 1-35 shows results of the comprehensive study by Coley and Heelis (1998), using DE in-situ satellite at about 840 km. The occurrence frequency peaks during winter, around UT ~ 12 to 24. Patch occurrence is suppressed in summer because the daytime sunlit polar-cap electron densities are already large, and difficult to double by transporting plasma from elsewhere. The UT dependence is due to the magnetic pole displacement (about 10 degrees in the northern hemisphere) from the earth's axis of rotation. The plasma flow lines in Figure 1-34 are in the geomagnetic latitude frame of reference, reflecting transpolar plasma transport bringing plasma into the polar cap through the cusp region, near 75 degrees magnetic latitude near noon. The reservoir of high-density sunlit plasma is in geographic coordinates, reflecting solar EUV plasma production. The magnetic pole (and cusp entry region) then dip 10 degrees more deeply into lower-latitude sunlit plasma about Canada, and 10 degrees farther from sunlit plasma about Siberia. For favored UTs, the magnetic entry region tips more deeply into lower-latitude sunlit high density plasma; for unfavored UTs, it draws from dark low-density plasma. This difference largely explains the northern hemisphere UT dependence of patch occurrence. The magnetic pole offset is not symmetric north to south pole, nor is the patch frequency of occurrence. As the auroral oval expands (or contracts) sunlit plasma will be drawn upon for a longer (or shorter) fraction of the day.
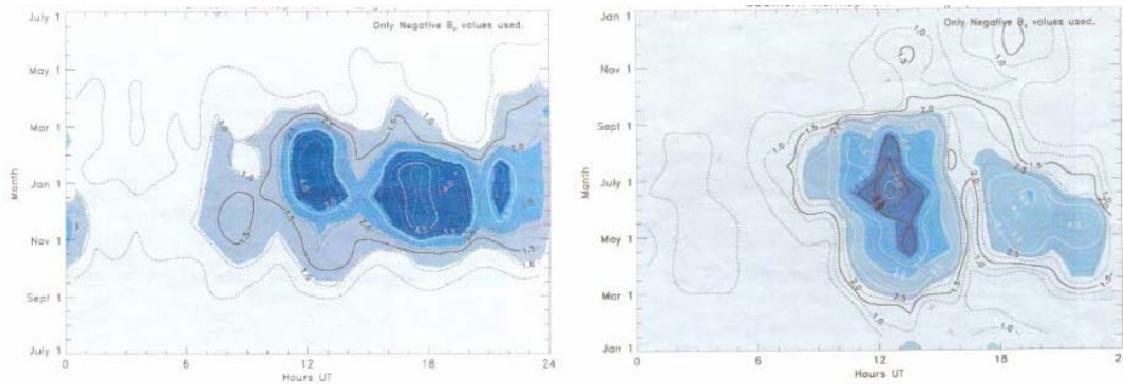
Figure 1-35: LH Side, Contour Plot of Occurrence Frequency of Polar Patches as Measured by DE 2 in the Northern Hemisphere vs. UT and Month of the Year, when the z Component of the IMF is Negative. The units on the contour lines are the number of patches observed per hour of spacecraft observation above 75 ILAT (from Coley and Heelis, 1998; Basu and Valladares, 1999.) RH side same but for southern hemisphere, showing strong hemispheric asymmetry.

Source Reservoir: Solar EUV-produced plasma density in this subauroral reservoir has well-documented seasonal and solar cycle variations, which reflect in variations in peak densities drawn into the polar cap. At times, events significantly enhance the plasma densities in the reservoir, and in its path between its region of solar EUV production and the cusp. These events can have very dramatic temporal/longitudinal consequences for the reservoir (Foster et al., 2005) that trace all the way across the polar cap. In the last decade or so, major progress has been made in understanding such processes, in the observing networks to track them, and in modeling them. Afternoon to pre-midnight electron densities at the equatorward edge of the trough can become significantly enhanced early in magnetic storms, a phenomena referred to as storm-enhanced density (SED) (Foster, 1993). This enhanced plasma is then transported towards noon by electric fields referred to as subauroral polarization streams (SAPS). This westward-flowing plasma, which can have major density increase through downward drainage of plasmaspheric plasma into the ionosphere, is not only of major interest to global ionospheric behavior, but can dramatically increase patch densities (Foster et al., 2005). We distinguish here between the plasma reservoir, and the mechanism that draws upon this reservoir to form patches, by pulling them into the polar cap in segmented islands. The physics of the subauroral plasma reservoir is treated in more detail elsewhere in this document.

When the magnetic pole tips towards the sun, so the magnetic cusp entry region continuously taps high-density plasma, one does not get a continuous smooth tongue of high-density plasma crossing the polar cap. To the contrary, the plasma is segmented, and the segments are severely structured, where the segmentation process is likely critical to the structuring. What leads to this segmentation?

Mechanisms for Patch Production: While there are no doubt many ways in which polar-cap patches can be formed, and at least six mechanisms have been published, there is no agreement as to which one(s) are most important or dominate (Lockwood and Carlson, 1992; Sojka et al., 1993; Rodger et al., 1994; Valladares et al., 1994, 1998).

Mechanisms proposed include, discrete changes in IMF B_y, B_z, and solar wind density/speed/pressure; time variation of average-flow-pattern models (e.g., Anderson et al., 1988); plasma flow jets channels in which relative-velocity-dependent recombination rates cut continuous tongues of plasma into segments and IMF reversals (Valladares et al., 1998); plasma production by cusp particle precipitation (Rodger et al., 1994, Millward et al., 1999); and transient magnetopause reconnection (Lockwood and Carlson, 1992), followed by Alfven wave coupling (Prikryl et al, 1999) of the solar wind to the ionosphere. The problem is not how patches might be produced, because a discrete change in many parameters can do that. The real question needs to be, how are most observed patches produced?

Cusp particle precipitation can produce patches up to ~ 5MHz f_oF_2, but models show it cannot explain the common 9 MHz patches, which produce severe scintillation of most concern. The increased recombination in high-speed flow channels can, in principle, segment tongues of high-density plasma to produce patches. However, this mechanism does not work, at least in the European sector, where high-speed flow channels produce higher plasma densities, not lower. The chemistry applies only where neutral densities are high enough to give significant recombination; it applies to plasma near and below 250 km, not much above. Anderson et al. (1988) have modeled patches by switching between different average large-scale flow patterns, but most patches appear to derive from transient events. Finding the dominant patch production mechanism, though challenging, is of more than academic interest. The production mechanism defines signatures for early identification, early tracking of direction, and even where in the polar cap scintillation-producing irregularities set in (Moen et al., 2001).

Recently one mechanism, the transient magnetopause reconnection mechanism, has been proven to work, to create patches, and furthermore to be a dominant patch-generation mechanism. Carlson et al. (2006), using the Svalbard EISCAT Radar (ESR) with supporting ASIP data, showed detailed agreement between model prediction and observations during an extended series of patches within the polar cap, observed coming directly from a subauroral plasma reservoir. The mechanism is a transient poleward excursion of the boundary between co-rotating and poleward moving plasma. The boundary then relaxes poleward. Observations confirmed the theoretical prediction of detailed coincidence in time and space of six plasma parameter boundaries, including optical arc boundaries that established motion of the patch boundary at formation speeds of ~2km/s. These observations establish clear patch onset signatures in parameters ideal for remote detection and early warning of patch onset.

Plasma Structuring and Scintillation: Keskinen and Ossakow (1983) have developed a well-tested theory for high-latitude ionospheric F-region irregularities (Tsunoda, 1988). This theory has led to the common acceptance of gradient drift instability as the mechanism that produces scintillation in mature patches. A further significant advance of experiment and theory has been the differentiation of this gradient drift instability and the velocity shear instability, which produces scintillation in sun-aligned arcs (Basu et al., 1998). These two plasma instabilities, which are triggered under very different conditions, were separated by in situ satellite measurements showing that the fractional fluctuation in electron density varied with fractional fluctuation in associated electric field, with tenfold more sensitivity in the velocity shear than in the gradient instability. Further work remains in theory and modeling of capacitive coupling to the magnetosphere above, coupling to variable conductivity below, and onset times. Also, irregularities in patches, attributed to the gradient drift instability, are observed earlier in the polar cap than is easily explained. While many patch-generating mechanisms leave this dilemma unresolved, the transient magnetopause mechanism for patch production would explain production of a patch filled with scintillation-producing irregularities by the process of patch creation.

Transpolar Trajectory, Exit, and Return Flow: Entry to the polar cap will differ significantly in initial velocity and plasma stress from one production mechanism to another. For example, magnetopause reconnection has initial >~2 km/s strong east or west B_y-dependent motion, compared to slower entry and weaker east-west motion for some other mechanisms. Once inside the polar cap, though, patches are assumed to follow well-established IMF-dependent transpolar flow paths (Heelis, 1984). Real time data can significantly improve accuracy and detail of large scale plasma flow (Ruohoniemi, 1993). Patch entry and exit have been modeled using time-varying patterns of flow, for flow patterns derived from averaged flow. Whether the exit of plasma through the nightside aurora is smooth or transient has been questioned.

"Boundary blobs" is the name given to enhanced ionospheric F-region plasma densities that are observed moving sunward near the equatorward edge of the auroral oval, primarily in the late afternoon and evening hours. These so-called blobs also represent walls of irregularities that produce scatter and scintillation. They were presumed to be patches that had exited from the midnight sector of the auroral oval. This assumption has not been confirmed, leaving their origin unproven. Their assumed exit from the polar cap, distortion of

shape, and return sunward flow has been tested once by Anderson et al. (1996). That work successfully models one representative patch example and identifies remaining issues for other scenarios.

1.3.7.5   Sun-Aligned Arcs and Theta Aurora: Theory and Modeling Context

The morphology of sun-aligned arcs has been well described by Valladares et al. (1994). That work describes frequency and location of IMF-dependent occurrence, orientation, motion, and other morphological properties. Much has also been learned about the near-earth electrodynamics, thermal structure, and energetics of these arcs (Valladares and Carlson, 1991). Arcs and patches have been observed in the same field of view after an IMF reversal, but only one from south to north, never from north to south, consistent with theoretical expectation that patches had been formed within the hour before the IMF reversal and had not had time to exit before the rapidly-forming arcs entered the field of view. Morphology of Theta aurora has been well described by Frank et al. (1986), within the context of associated particle energy distributions, currents, and the full suite of in situ measurements on board the DE satellite. Many characteristics mimicked those of sun-aligned arcs, but all were less dramatic. The number of sun-aligned arc events decreases with increasing intensity, and an intensity- vs. frequency-of-occurrence plot would put the intensity of Theta aurora at about their observed frequency of occurrence. Therefore, morphology alone leads us to ask, are they fundamentally the same, or fundamentally different? Distinctions will be discussed later below, between common, weak sun-aligned arcs and rare, intense Theta aurora, within the context of modeling and theory. Figure 1-36 shows representative sun-aligned arcs and Theta aurora.
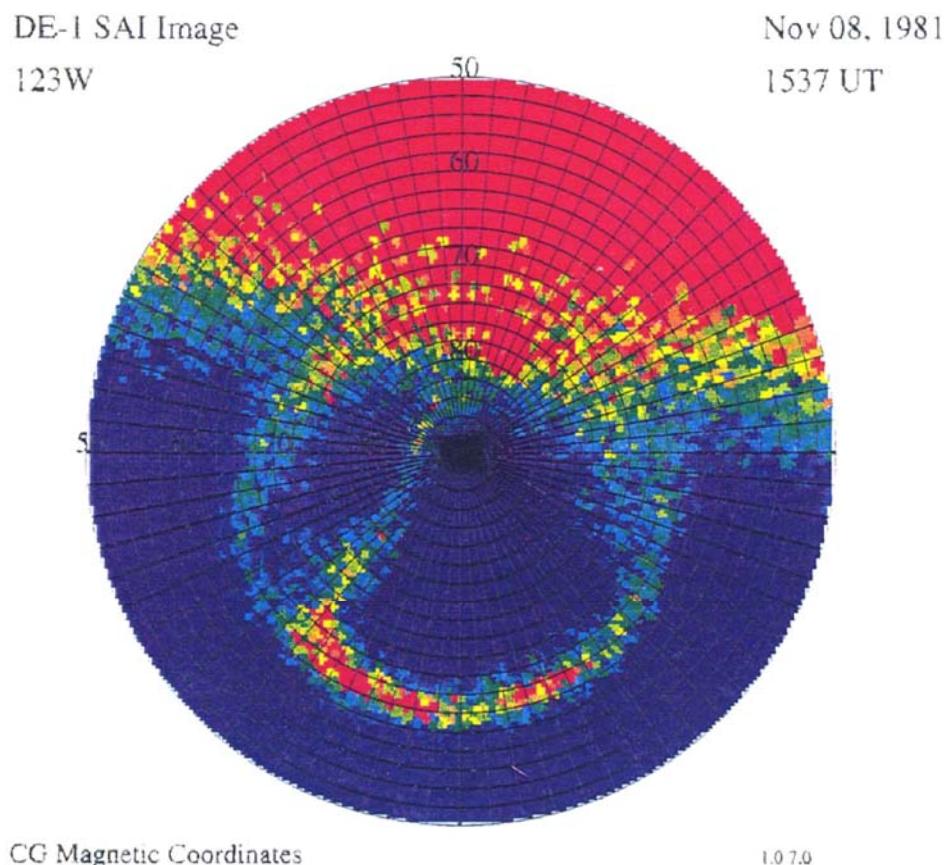
Figure 1-36a: A Theta Aurora Observed by the DE Satellite, Spanning the Polar Cap from the Midnight to the Dayside Auroral Oval, or Cusp, Near the Earth Limb (Frank et al., 1986). Theta aurora are seen a few percent of the time, and only after extended northward IMF.
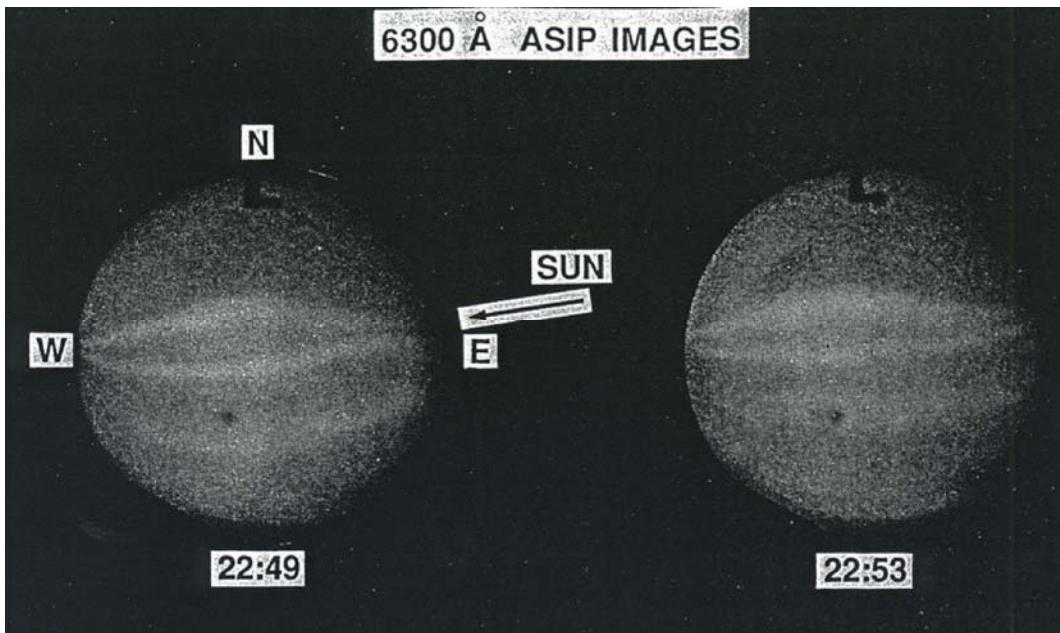
Figure 1-36b: Sun-Aligned Arcs in 1000 km Field of View are Present Nearly Half the Time, that Half when the IMF is Northward. The black dot is a fixed point on the ground, with scintillation-free field of view overhead at 22:40 UT and strong scintillation at 22:53 UT. Note the difference in both field of view, and spatial resolution, between this ground-based SIP and satellite-borne platforms (Valladares et al. 1994).

Outstanding theoretical and modeling issues have been the mapping of boundaries between the ionosphere and magnetosphere, the direction of drift and symmetry/asymmetry between northern and southern hemisphere (Obara et al. 1988), particle sources for current sheets, the fundamental mechanism for formation, and what, if any, is the fundamental difference between sun-aligned arcs and Theta aurora. Resolution of these questions is key to moving on to a predictive modeling capability and quantitative understanding.

1.3.7.5.1   Mapping to the Magnetosphere

Studies of data from satellite-borne particle detectors, sometimes including coincident optical imagery from satellites or the ground, have led to seemingly contradictory interpretations or conclusions about where polar-cap arcs map to in the magnetosphere. Examples are the 12 "cartoon" models shown in the figures described in section 6.2.3 of the book by Sandholt et al., 2002. The sun-aligned arcs have been either directly observed with coincident images, or inferred to have been present on the basis of low-earth-orbit satellite particle measurements. The different magnetospheric regions to which sun-aligned arcs have thus been proposed to connect include, closed field lines threading an expanded (e.g., Meng, 1981) or bifurcated (e.g., Frank et al., 1986) plasma sheet and associated with sunward flow; open field lines linking the mantle and magnetosheath (e.g., Hardy et al., 1982), or closed field lines mapping to the lower-latitude boundary layer (LLBL) and associated with antisunward flow (e.g., Lundin et al., 1991). Weiss et al., (1993), have schematically mapped from the high-latitude ionosphere to the distant magnetotail for five models of the polar-cap morphology, namely open, bifurcated, expanded plasma sheet, expanded LLBL, and rotated X line. There is no consensus on the mapping to the magnetosphere.

1.3.7.5.2   Drift

The sun-aligned arcs are observed to drift sideways, at ~0.1 km s^{-1}. The direction of dawn-dusk drift of the sun-aligned arcs, duskward (dawnward) for $B_y > 0$ ($B_y < 0$), has been suggested as being due to dayside

merging (Valladares et al., 1994; Newell and Meng, 1995). Many models are based on Crooker's antiparallel merging hypothesis (Crooker, 1979). Reconnection topology presented by Reiff and Burch (1985) can drive convection cells by merging in the cusps, where, if the magnetotail field lines are open, they recover the single-lobe cell of Crooker (1979) with B_y-dependent sense of rotation.

1.3.7.5.3   Currents

The near-earth ionospheric behavior is much better understood. The generation of large-scale regions of auroral currents and particle precipitation by the divergence of convection electric fields was treated in detail by, and explicitly applied to the mesoscale electrodynamics of observed transpolar and sun-aligned arc segments by Carlson et al. (1988). The essential electrodynamic, thermal, and energetic properties of sun-aligned arcs at ionospheric altitudes were calculated by Valladares and Carlson (1991). Sun-aligned arcs exist within plasma velocity gradients of a sense which produces converging electric fields. In the northern polar cap, antisunward plasma flow velocity decreases across the arc in the dawn to dusk direction, with this gradient sometimes observed to extend into sunward flow on the dusk edge. The particle precipitation and consequent optical emission ends where the gradient reverses or ends, as for an Ohm's law arc. The arcs are called sun-aligned arcs because the flow lines are strongly anisotropic, with scale sizes >1000 km along the sun-earth line, while only approximately tens to hundreds of kilometers in the dawn-dusk direction.

1.3.7.5.4   Formation

Theta auroras are sun-aligned arcs which have intense emissions, well exceeding a kR in 630.0 nm emission, the electrons exciting them having relatively high energy, exceeding a keV. They connect fully from the day to nightside auroral oval, and are seen to drift dawn to dusk as do weaker, common, quickly-forming sun-aligned arcs. Theta aurora are seen only after many hours of uninterrupted northward IMF, and are best, if not only, seen from satellites. They are rare, being seen only a few percent of the time, but they are dramatic and have attracted considerable attention in the literature since first noted (Frank et al., 1986). These authors, on the basis of the event they reported on, argued that the theta aurora was located on closed field lines. They argued for a polar cap consisting of two open cells bifurcated by a closed region mapping to the plasma sheet boundary layer or the distant plasma sheet.

For a recent sketch of the large number (nine) of models devoted to explaining the formation of Theta aurora, their conflicting predicted behavior, and consistency with data (e.g. Carlson and Cowley, 2005).

For common sun-aligned arcs, Carlson and Cowley (2005) applied a kinetic theory perspective of electrons precipitating into the polar cap from the electron reservoir in the distant magnetospheric tail, for currents sustainable without electron acceleration. They concluded that for typical sun-aligned arc conditions, there is no need to look further than the magnetospheric tail for the source reservoir of electrons needed to provide observed fluxes in either polar rain or sun-aligned arcs. Applying this theory to this ready reservoir, they suggested that the sun-aligned arcs are indeed produced by field-aligned acceleration of polar rain electrons. The implied field-aligned voltages are ~200–600 V, with inferred minimum heights of the acceleration region of ~0.7–0.9 R_E above the polar ionosphere. Finding the likely source of the electrons drawn upon to produce sun-aligned arcs should be seen as a significant step toward the resolution of the long- question standing about the topology of the magnetosphere under northward IMF conditions.

Recent work (Milan et al., 2005 and references therein) has turned more serious effort towards sun-aligned arc formation, using concepts such as those of Cowley and Lockwood (1992) and Reiff and Burch (1985), at a high-latitude or lobe-reconnection site, as expected when the IMF is northward. It is expected that this approach will lead to new insights in the near future, regarding the ubiquitous sun-aligned arcs.

1.3.7.5.5   Contrasting Formation of Common Sun-Aligned vs. Theta Aurora

Interpretation of many observations of Theta aurora particle properties indicate that plasma sheet precipitation on closed field lines is the cause of Theta aurora. Theory and modeling support this conclusion. Theta aurora are the class of polar-cap sun-aligned arcs that are more than a kR in intensity (and thus bright enough to be visible to the naked eye), and which result after hours of northward IMF conditions, affording enough time for significant magnetospheric topological reconfiguration. The location of theta aurora on closed field lines suggests that the magnetospheric dynamics has resulted in a bifurcation (Frank et al., 1986) of the tail lobe, due, for example, to a reversal in B_y during the interval which takes time to develop. Overall, theta auroras are found in the polar cap a few percent of the time.

In sharp contrast, sun-aligned arcs are present in the polar cap nearly half the time, that half of the time that the IMF is northward. These arcs are sub-visual intensity, typically hundreds of R. They have received less attention in the literature, perhaps because they are difficult to observe from satellites, and are less dramatic in commonly available images. However, they define the dominant character of the northward IMF polar cap, including its highly anisotropic shear flow character (Carlson et al., 1988). The ubiquitous sub-visual sun-aligned arcs are consistent with accelerated polar rain on open field lines. These arcs can be generated by any modest mechanism that drives a flow shear across the open field lines. Such mechanisms include, for example, the boundaries between lobe cells and merging cells (open flux production does not switch off completely until the IMF is quite close to pure north, perhaps within ~30–40 degrees), or flows from differing magnetopause reconnection sites. Such conditions can easily occur. Any such process giving rise to flow shears on open field lines, leading to upward field-aligned currents bigger than 0.01 mA m^{-2}, will require acceleration, and will produce polar shower precipitation and a corresponding aurora.

In short, present theory indicates that the rare Theta aurora are on closed field lines, and the ubiquitous sun-aligned arcs are on open field lines.

1.3.7.6   Future Work

Our understanding today of polar-cap phenomena includes a detailed characterization, much morphology, and definition of many of the processes determining behavior. Consequently, we have now developed valuable techniques to mitigate the effects of disruptive polar phenomena on high-latitude communication and GPS navigation systems. However, only a physics-based model, tested and validated, can deliver a predictive capability that will allow uninterrupted connectivity where needed. Much work remains to achieve this.

We do not have agreed-upon physical models of the entry, transit, exit, or return flow of the patches characteristic of southward IMF conditions. Neither do we have agreed-upon physical models for rare Theta aurora or the sun-aligned arcs characteristic of northward IMF conditions. We have much to discover about the processes that couple the polar ionosphere to the magnetosphere, and to the solar wind looking outward towards far space, and about how polar-cap coupling impacts the dynamics of the global ionosphere, our near-space earth environment. The mechanism for the creation of polar-cap patches has been debated for 20 years now, without a consensus. The one validated mechanism has limited early-warning time, and we are unclear on the significance of other candidate mechanisms. Accurate prediction of patch trajectory across the polar cap is key to the timing and location of downstream outages of 10 minutes or more. Yet transpolar models are statistical, and their application to patches exiting through the nightside aurora and returning with a westward flow is questionable. Conflicting models exist for Theta aurora and sun-aligned arc formation, with conflicting predicted behavior. We must even ask ourselves, to what degree is polar plasma large-scale flow dominated by large-scale fields rather than being the sum of many small-scale transient reconnection events.

Recent advances in experimental and instrumental capability, and computer modeling, make this a time ripe for a surge of progress. Recent ground-based measurements finally have the temporal and spatial

resolution to diagnose the evolution of key transient events, rather than only seeing before-and-after states. Satellite constellations can now sample fields and particles at sites of critical phenomena, with the multipoint ability to separate temporal from spatial structure. Computer power now dissects complex interactive processes, and focuses instrument development and experiment deployment on the most critical challenges. It is not unreasonable to suggest that the next decade might even reveal the nature of magnetic reconnection itself, a discovery of fundamental importance to plasmas from laboratory to galactic scales. Strong dialogue between the radio and space science communities will compound the gains.

1.3.7.7 Caveat

There is a great amount of on-going progress, from the work of a large and vigorous research community, in the area of observing, modeling, and understanding the character and space physics of the polar-cap ionosphere. In a sketch as short as this, it is inevitable that much excellent and important work could not be covered, in some cases not even noted. The intent is to present to radio professionals concerned with phenomena and effects of the polar-cap ionosphere, a framework, and enough information to pursue those areas of most concern to them.

1.3.7.8 Acknowledgements

1.3.7.9 References

Anderson, D. N., J. Buchau, and R. A. Heelis, Origin of density enhancements in the winter polar cap ionosphere, *Radio Sci., 23*, 513-519, 1988.

Anderson, D. N., D. T. Decker and C. E. Valladares, Modeling Boundary Blobs using time varying convection, *Geophys. Res. Lett., 23*, 579-582, 1996.

Basu, S., E. MacKenzie, and Sunanda Basu, Ionospheric constraints on VHF/UHF communications links during solar maximum and minimum periods, *Radio Sci., 23*, 363-378, 1988.

Basu, S. and C. E. Valladares, Global aspects of plasma structures, *J. Atmos. Solar-Terr. Phys., 61*, 127, 1999.

Basu, S., S. Basu,, E. MacKenzie, P. F. Fougere, W. R. Coley, N. C. Maynard, J. D. Winningham, M. Sugiura, W. B. Hanson and W. R. Hoegy, Simultaneous density and electric field fluctuation spectra associated with velocity shears in the auroral oval, *J. Geophys. Res*. *93*, 115-136, 1998.

Buchau J., et al., Ionospheric structures in the polar cap: Their origin and relation to 250-MHz scintillation, *Radio Sci*., *20*, 325-338, 1985.

Buchau, J., and B. W. Reinish, Electron density structures in the polar F-region, *Adv. Space Res*., *11* (10), 29, 1991.

Carlson, H. C., R. A. Heelis, E. J. Weber, and J. R. Sharber, coherent mesoscale convection patterns during northward IMF, *J. Geophys. Res*., *93*, 14,501-14,514, 1988.

Carlson, H.C., The dark polar ionosphere: Progress and future challenges, *Radio Sci*. *29*, 157-165, 1994.

Carlson, H. C., J. Moen, K. Oksavik, C. Nielsen, I. McCrea, T. Pedersen, P. Gallop, Direct observations of injection events of subauroral plasma into the polar cap, *Geophys. Res. Lett, 33*. L05103, 2006.

Carlson, H. C. and S. W. H. Cowley, Accelerated polar rain electrons as the source of sun-aligned arcs in the polar cap during northward IMF conditions, *J. Geophys. Res., 110*, A05302 (doi: 10.1029/ 2004JA010669), 2005.

Coley, W. R. and R. A. Heelis, Structure and occurrence of polar ionization patches, *J. Geophys. Res., 103*, 2201, 1998.

Cowley, S. W. H., and M. Lockwood, Excitation and decay of solar wind-driven flows in the magnetosphere-ionosphere system, *Ann. Geophys, 10*, 103, 1992.

Crooker, N.R., Dayside merging and cusp geometry, *J. Geophys. Res., 84*, 951, 1979.

Crowley, G., A critical review of ionospheric patches and blobs, *Rev of Radio Science 1993-96*, Oxford Univ. Press, 619-648, 1996.

Dandekar, B. S., and T. W. Bullett, Morphology of polar-cap patch activity, *Radio Sci., 34*, 1187, 1999.

Foster, J. C., Storm time plasma transport at middle and high latitudes, *J. Geophys. Res., 98*, 1675, 1993.

Foster, J. C., et al, Multiradar observations of the polar tongue of ionization, *J. Geophys. Res., 110*, A09S31, 2005.

Frank, L., et al., The theta aurora, *J. Geophys. Res., 91*, 3177, 1986.

Gussenhoven, M. S., Extremely high latitude auroras, *J. Geophys. Res., 87*, 2401, 1982.

Hardy, D. A., W. J. Burke, and M.S. Gussenhoven, DMSP optical and electron measurements in the vicinity of polar-cap arcs, *J. Geophys. Res., 87*, 2413, 1982.

Heelis, R.A., The effects of interplanetary magnetic field orientation on dayside high-latitude convection, *J. Geophys. Res. 89*, 2873, 1984.

Ismail, S., D. D. Wallis, and L. L. Cogger, Characteristics of polar-cap Sun-aligned arcs, *J. Geophys. Res. 82*, 4741, 1977.

Keskinen, M. J., and S. I.. Ossakow, Theories of high latitude ionospheric irregularities, *Radio Sci, 18*, 1077, 1983.

Lockwood, M. and H. C. Carlson, Production of polar cap electron density patches by transient magnetopause reconnection, *Geophys. Res. Lett., 19*, 1731, 1992.

Lundin, R., L. Eliasson, and J. S. Murphree, The quiet-time aurora and the magnetosphere configuration, in Auroral Physics, edited by C.-I. Meng, M. J. Rycroft, and L.A. Frank, p. 177, Cambridge Univ. Press, New York, 1991.

McEwen, D. J., and D. P. Harris, Occurrence patterns of F region layer patches over the north magnetic pole, *Radio Sci., 31*, 619-628, 1996.

Meng, C.-I., Polar-cap arcs and the plasma sheet, *Geophys. Res. Lett., 8*, 273, 1981.

Milan, S. E., B. Hubert and A. Grocott, Formation and motion of a transpolar arc in response to dayside and nightside reconnection, *J. Geophys. Res., 110*, A01212, 2005.

Millward, C. H., et al., Modeling the ionospheric effects of ion and electron precipitation in the cusp, *J. Geophys. Res., 104*, 24603. 1999.

Moen et al, On the collocation between dayside auroral activity and HF Radar backscatter, *Ann. Geophys., 18*, 1531-1549, 2001.

Newell, P.T., and C.-I Meng, Creation of theta-auroras: The isolation of plasma sheet fragments in the polar cap, *Science, 270*, 1338, 1995.

Obara, T., M. Kitayama, T. Mukai, N. Kaya, J. S. Murphree and L. L. Cogger, Simultaneous observations of Sun-aligned polar-cap arcs in both hemispheres by EXOS-C and Viking, *Geophys Res. Lett., 15*, 713, 1988.

Prikryl, P., J. W. MacDougal, I. F. Grant, D. P. Steel, G. J. Sofko, and R. A. Greenwald, Observations of polar patches generated by solar wind Alfven wave coupling to the Datside magnetosphere, *Ann. Geophysicae, 17*, 463-489, 1999.

Reiff, P.H. and J. L. Burch, IMF B_y dependent on dayside plasma flow and Birkeland currents in the dayside magnetosphere 2: A global model for northward and southward IMF, *J. Geophys. Res., 90.* 1595, 1985.

Rodger, A. S., et al., A new mechanism for polar patch formation, *J. Geophys. Res., 99*, 6425, 1994.

Ruohoniemi, J. M. and K. B. Baker, Large-scale imaging of high latitude convection with Super Dual Auroral Radar Network HF radar observations, *J. Geophys. Res., 103*, 20797, 1993.

Sandholt. P. E., H. C. Carlson, and E. Egeland, *Dayside and Polar-cap Aurora*, Kluwer Academic Publishers, Dordrecht Netherlands, 2002.

Schunk, R. W., and J. J. Sojka, Ionosphere-thermosphere space weather issues, *J. Atmos. Terr. Phys., 58*, 1527, 1996.

Smith, M. F., and M. Lockwood, Earth's Magnetospheric Cusps, *Rev. Geophys., 34*, 233-260, 1996.

Sojka, J. J., et al., Modeling polar cap F-region patches using time varying convection, *Geophys. Res. Lett., 20*, 1783, 1993.

Sojka, J. J., M. D. Bowline and R. W. Schunk, Patches in the polar-cap ionosphere: UT and seasonal dependence, *J. Geophys. Res., 99*, 959, 1994.

Tsunoda, R. T., High-latitude F region irregularities: a review and synthesis, *Rev. Geophys 26*, 719, 1988.

Valladares, C.E. and H, C, Carlson, The electrodynamic thermal and energetic character of intense sun-aligned arcs in the polar cap, *J. Geophys. Res.* 96, 1379-1400, 1991.

Valladares, C. E., H. C. Carlson and K. Fukui, Interplanetary magnetic field dependency of stable sun-aligned polar-cap arcs, *J. Geophys. Res., 99*, 6247-6272, 1994.

Valladares, C. E., et al., Formation of polar cap patches associated with north-to-south transitions of the IMF, *J. Geophys. Res., 103*, 14,657, 1998.

Vickrey, J. F., C. L. Rino, and T. A. Potemra, Chatanika, TRIAD observations of unstable ionization enhancements in the auroral F-region, *Geophys. Res. Lett.* 7, 789-792, 1988.

Weber, E.J. and J. Buchau, Polar-cap F-layer auroras, *Geophys. Res Lett. 8*, 125, 1981.

Weber et al., F-layer ionization patches in the polar cap, *J. Geophys. Res., 89*, 1683, 1984.

Weber, E.J., J. A. Klobuchar, J. Buchau, H. C. Carlson, R. C. Livingston, O. de la Beaujardiere, M. McCready, J. G. Moore and G. J. Bishop, Polar cap it F layer patches: structure and dynamics, *J. Geophys. Res. 91*, 121, 1986.

Weiss, L.A., E. J. Weber, P.H. Reiff, J. L. Sharber, J. D. Winningham, F. Primdahl, I. S. Mikkelsen, C. Seifring, and E. M. Wescott, Convection and electrodynamic signatures in the vicinity of a Sun-aligned arc: Results from the polar acceleration regions are convection study (polar arcs), in *Auroral Plasma Dynamics, Geophys. Mongr. Ser., 80* edited by R. L. Lysak, p. 69, AGU, Washington, DC, 1993.

Zhu, L., R. W. Schunk, and J. J. Sojka (1997), Polar-cap Arcs: A review, J. Atmos. Solar-Terr Phys., 59, 1087, 1997.

1.3.8 Traveling Ionospheric Disturbances (TIDs)

(Paul Prikryl, Communications Research Centre Canada)

1.3.8.1 Relationship between Atmospheric Gravity Waves and TIDs

Atmospheric gravity (buoyancy) waves (AGWs) with sources in the lower and upper atmosphere, particularly those originating in the auroral electrojets, produce traveling ionospheric disturbances (TIDs). As the gravity wave propagates in the thermosphere the ionosphere is structured into quasi-periodic plasma density enhancements and depletions moving with the horizontal phase velocity of the wave. The neutral motions due to the gravity wave are transferred to ions and electrons through collisions (ion drag). Because the collision rate depends on the altitude, the ionospheric response to gravity waves is altitude dependent. In the F-region, where the collision frequency is small compared to the gyrofrequency, the ion motion perpendicular to the magnetic field is inhibited. The ions move parallel to the geomagnetic field, which has a tilt from the vertical that increases with decreasing latitude. Although the effect of a passing gravity wave is manifested in the ionosphere as a TID, the observed F-region motions and densities of ions, which are coupled to electrons, by no means precisely represent the gravity-wave-induced neutral motions and densities. Nevertheless, some of the wave parameters, namely wavelength and period, can be inferred from observations of TIDs (Kirchengast, 1996).

In general, AGW/TIDs are divided into two broad classes according to their phase velocity and period, namely large- and medium-scale waves (see, e.g., Hocke and Schlegel, 1996). The large-scale waves propagate in the thermosphere with horizontal phase velocities between 400 and 1000 m/s (comparable with the sound velocity in the thermosphere), have horizontal wavelengths greater than 1000 km and periods between about 1 and 3 h. The medium-scale waves, which may also propagate in the lower atmosphere before they are detected in the ionosphere, have horizontal phase velocities between 100 and 300 m/s (less than the sound velocity in the lower atmosphere), wavelengths of several hundred kilometers and periods between 15 and 60 min (Hunsucker, 1982).

1.3.8.2 Dispersion Relation and Ray Tracing of Gravity Waves

The theoretical basis for the understanding of AGWs and their role in the ionosphere was developed by Hines (1960). The gravity wave dispersion relation is consistent with both upward group (downward phase) and downward group (upward phase) propagation. Yeh and Liu (1974) reviewed propagation of AGWs in a realistic atmosphere using the ray theory approach and the WKB approximation. A simplified ray tracing procedure based on Snell's law is applicable for a horizontally stratified atmosphere. Assuming

a model atmosphere and a source of gravity waves in the auroral ionosphere, a simple two-dimensional ray tracing of the gravity waves can be conducted (Prikryl et al., 2005). Viscosity, thermal conductivity and neutral winds are neglected. Snell's law for a spherically stratified atmosphere with no horizontal gradients (Muldrew, 1959) is given by $nr \sin \alpha =$ constant where n is the refractive index, r is the distance from the centre of the earth to a point on the ray and α is the angle of incidence. The dispersion relation between the gravity wave frequency ω and wave vector **k** given by Hines (1960) is used:

$$(\omega^2 - \omega_a^2)\, \omega^2 / C^2 - \omega^2(k_x^2 + k_z^2) + \omega_b^2 k_x^2 = 0, \tag{1}$$

where $\omega_a = \gamma g/2C$ is the acoustic cutoff frequency, C denotes the speed of sound, and k_x and k_z are the horizontal and vertical components of the wave vector **k**. For a nonisothermal atmosphere (one in which the atmospheric scale height varies with altitude) the Brunt-Väisälä (buoyancy) frequency ω_b is defined as

$$\omega_b^2 = (\gamma-1)g^2/C^2 + (g/C^2)(dC^2/dz) \tag{2}$$

Figure 1-37 shows an example of group paths for gravity waves having a period of 30 min and originating at an altitude of 110 km, using the MSIS90 atmosphere model. The ray tracing is initiated at an altitude of 110 km using selected initial values α_0 of the **k** directions α, where α is taken to be positive upward from the southward horizontal direction in the northern hemisphere. The group velocity $\mathbf{V}_g \equiv (\partial\omega/\partial k_x, 0, \partial\omega/\partial k_z)$ or path of energy flow of the gravity wave is obtained from (1) numerically. The rays are the same in all panels but are colour-coded for (a) time of travel, (b) phase velocity V_x^{ph}, (c) horizontal wavelength λ_x and (d) vertical wavelength λ_z (Figure 1-37). For each ray, the wave vector directions are shown at the start of the ray tracing (α_0) and after one wave period. The ray tracing shows three distinct group paths that reach the F region, namely the rays that travel directly upward, rays that are reflected in the lower thermosphere, and rays that are reflected in the troposphere or from the earth's surface.
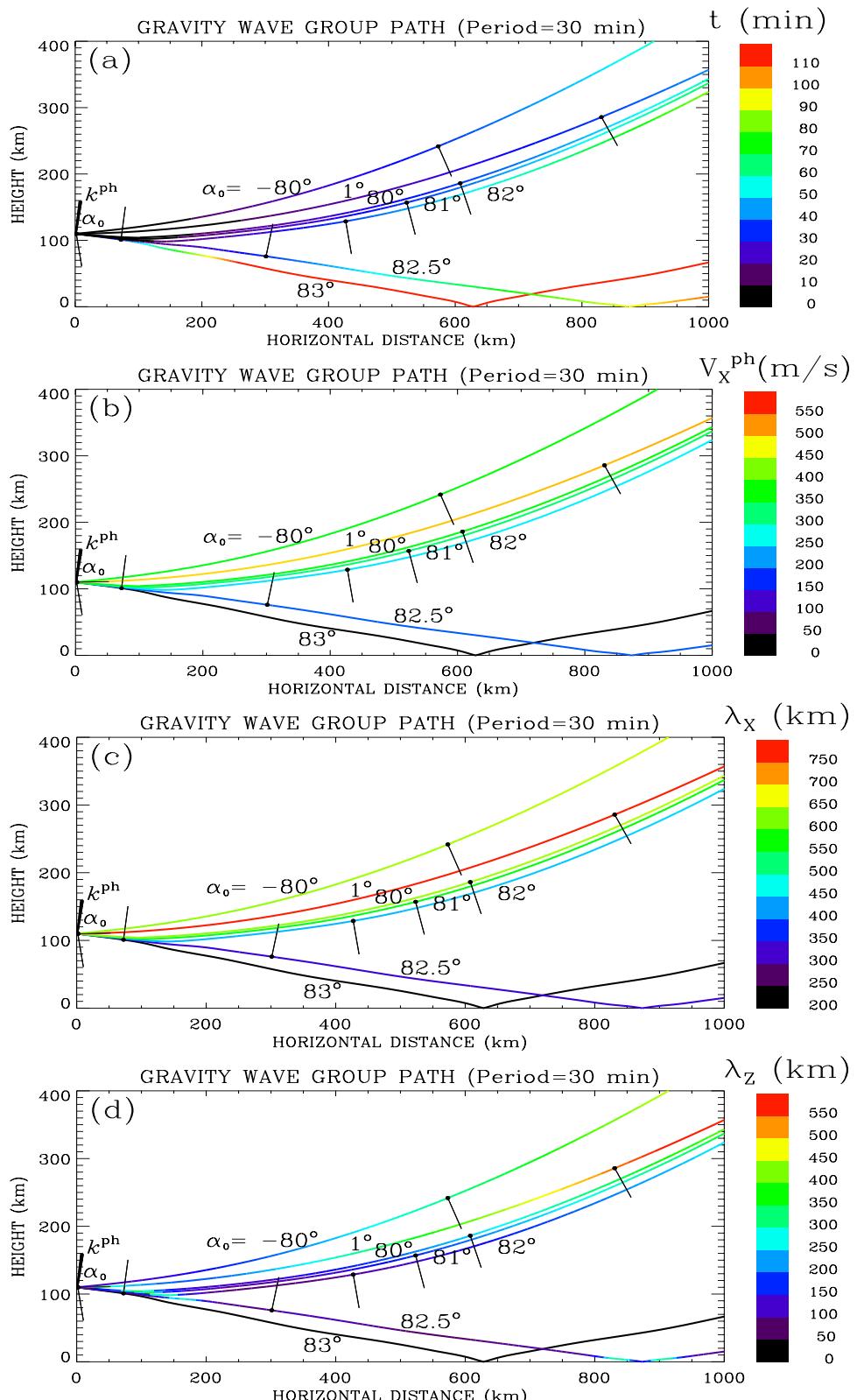
Figure 1-37: Ray Tracing of Gravity Waves (Group Path). When the phase propagation direction k^{ph} is upward ($\alpha_0 > 0$), the group velocity is downward and vice versa. The rays are the same in all panels, but are colour-coded for time of travel, horizontal wavelength and horizontal phase velocity (Prikryl et al., 2005).

1.3.8.3   Radio and Optical Techniques of AGW/TID Detection

AGW/TIDs can be observed using various techniques (see, e.g., Hunsucker, 1982; Davies, 1990; Hargreaves, 1992). In the ionosphere, the principal methods are based on radio techniques, namely ionosode, HF Doppler, HF radar (ground/sea scatter), incoherent scatter radar, total electron content (TEC) now routinely measured with GPS receivers and ionospheric tomography. Also, because the gravity waves modulate the night sky airglow (Mendillo et al., 1997), optical techniques provide a wealth of information on gravity waves.

a)   HF Doppler

A very sensitive method for detecting transient variations in the vertical motion of ionospheric layers is the HF Doppler technique operating in the 3-10 MHz band (Georges, 1968; Davies, 1990). Networks of HF Doppler systems were used to observe bottom-side F-region undulations due to TID/AGW and derive periods and horizontal phase velocities (see references in Lewis et al., 1996). The advantages of this technique over other techniques of TID measurement are relatively high temporal resolution (30 sec) and lower power consumption than traditional ionosonde chains.

b)   Ionosonde

Traditional ionosonde measurement of the virtual height displacement of the F region was traditionally used as indicator for large-scale TIDs. Changes in N_mF_2 and h_mF_2 can be detected. Extensive networks or chains of ionosondes have been deployed to study propagation of TIDs and to identify the sources of AGWs, particularly in the auroral electrojet, by measuring delays, phase velocities, wavelengths and amplitudes. Worldwide atmospheric gravity wave studies (WAGS) have been conducted in coordination with other instruments. Results from these studies that were summarized by Hocke and Schlegel (1996) provided a major improvement in understanding of TIDs. A modern digital ionosonde (see Chapter 2, and http://cadiweb.physics.uwo.ca) combines the capabilities of the HF Doppler and traditional ionosonde technique (Grant et al., 1995), thus allowing improved observations of TIDs (MacDougall et al., 1997).

c)   HF Radar

Radar backscatter sounding techniques for observation of motions of large-scale F-region irregularities date back to 1950s (Tveten, 1961; and references therein). Hunsucker and Tveten (1967) used the HF backscatter technique to observe large scale TIDs on ground-scatter records. At the ionospheric reflection point the signals are modulated (focused/defocused) by TIDs which are attributed to imperfectly-ducted atmospheric gravity waves generated by auroral sources. More recently, HF radars are used to observe ionospherically generated medium-scale gravity waves, which may be earth-reflected and then produce trains of TIDs many hundreds of kilometers equatorward of the original source region (Samson et al. 1989; 1990). These radars are now a part of SuperDARN (Super Dual Auroral Radar Network) (Greenwald et al., 1995; Chisham et al., 2007). SuperDARN has become a useful instrument for the study of TIDs (Bristow et al., 1994; Hall et al., 1999) because of the large spatial coverage. The combined radar field-of-view covers a large fraction of the auroral and polar region, and most recently, new radars have been installed in mid latitudes. Because SuperDARN was primarily designed to monitor ionospheric convection (Greenwald et al., 1995), the SuperDARN radars can observe the sources of AGW/TIDs, namely convection intensifications at high latitudes (Prikryl et at., 2005).

Figure 1-38 shows a sequence of the Kapuskasing radar maps of the received power showing the ground scatter signatures of two successive medium-scale TIDs. The ground scatter band 2 spanning the radar field of view moved equatorward as the next band 3 appeared about 400 km poleward of it. A series of ionospheric convection bursts at high latitudes caused by pulsed magnetic reconnection at the dayside magnetopause was the source of TIDs. A band of ionospheric scatter at far ranges (near

80° magnetic latitude) is the TID source region where the ionospheric signature of the reconnection, namely a convection pulse that is sometimes called a flow channel, occurred.
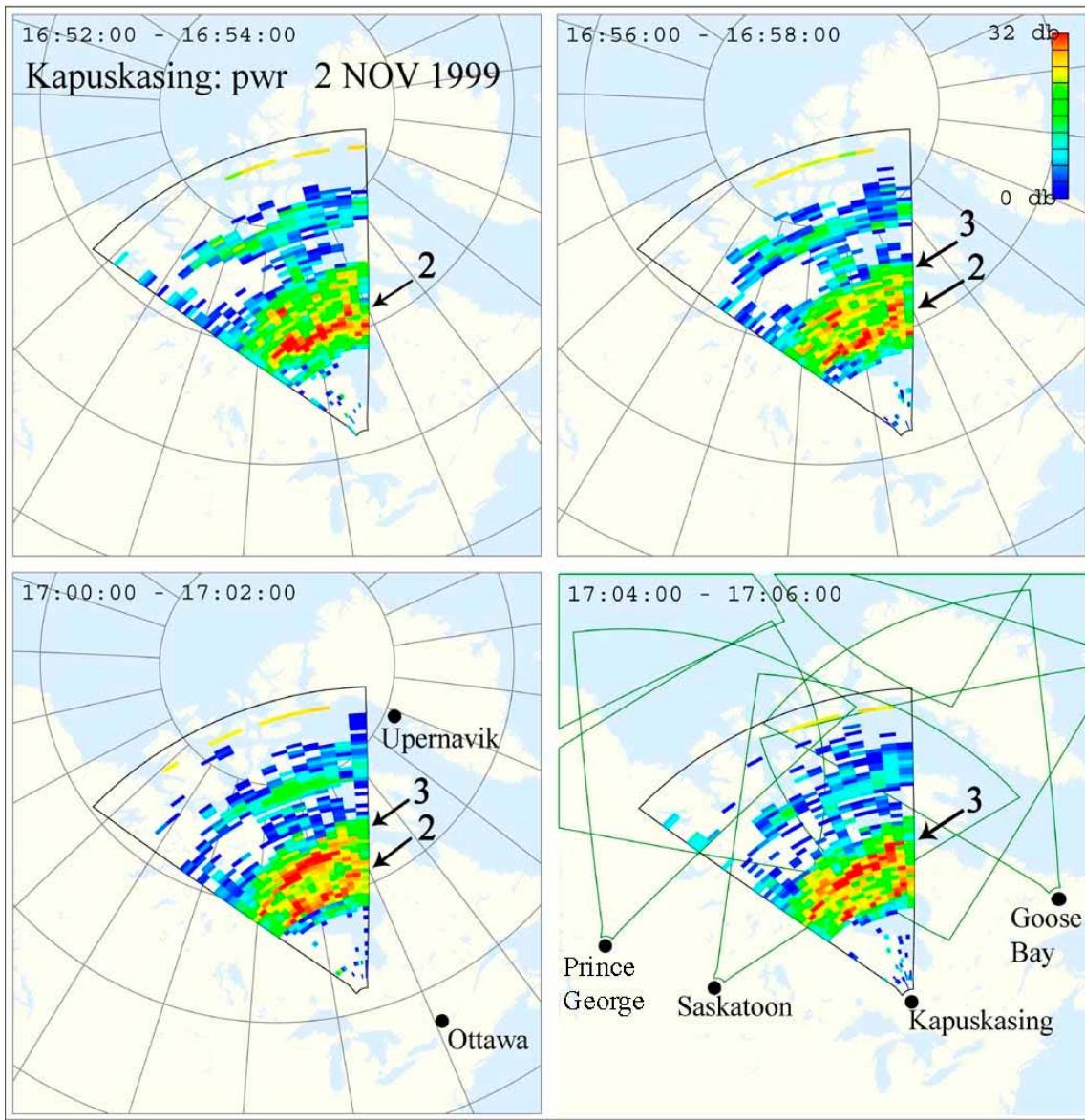
Figure 1-38: The Ground Scatter Signatures of Equatorward-Moving Medium-Scale TIDs in the Kapuskasing Radar Field-of-View.

Figure 1-39 shows the ionospheric convection/potential map constructed using the line-of-sight velocity data from all northern-hemisphere SuperDARN radars (Ruohoniemi and Baker, 1998). The fitted convection velocity vectors are shown colour-coded (see velocity scale bar top-left). The cross-polar-cap electric potential contour map is shown by solid and broken lines outlining the dusk and dawn convection cells, respectively. The value of the cross-polar-cap voltage is printed in the upper-left corner. Also, the IMF BZ versus BY clock-dial plot is shown (top-right). The convection map shows the ionospheric signature of a flux transfer event, namely an ionospheric flow channel extended over several hours of local time about 12 MLT.
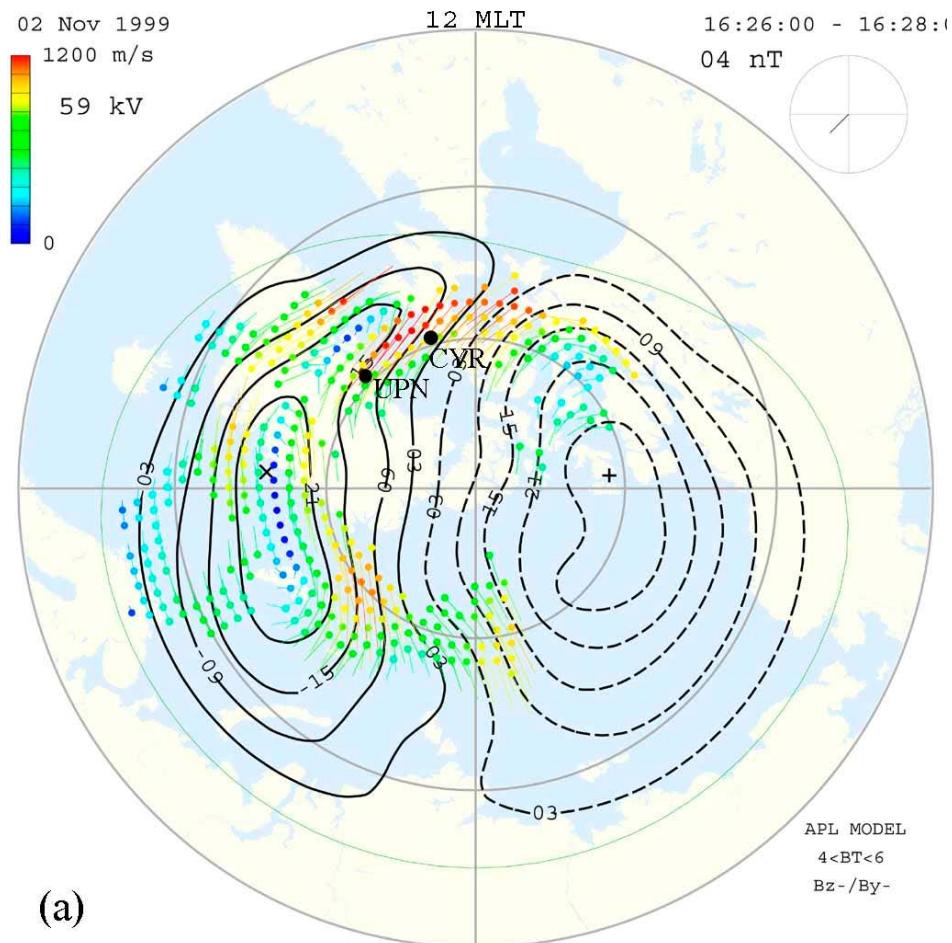
**Figure 1-39: The Ionospheric Convection Intensification Centered at 12 MLT
was the Source Region for the TID 2 shown in Figure 1-38.**

d) Incoherent Scatter Radar

Simultaneous measurements by incoherent scatter radars of various ionospheric parameters, such as electron density, ion velocity, and ion/electron temperature, have been used to retrieve TID parameters by means of inversion techniques (Hocke and Schlegel, 1996). To elucidate the physics of the AGW-TID relationship, theoretical simulations were conducted and compared with incoherent scatter radar (ISR) data (Kirchengast 1996; Kirchengast et al., 1996). Based on such theoretical investigations AWG characteristics may be derived or inferred from TID modelling in conjunction with incoherent scatter radar measurements (Kirchengast et al., 1995).

Large-scale TIDs have been extensively imaged using ISR techniques and later compared with tomographic reconstruction (Nygrén et al., 1996) and GPS measurements of total electron content (e.g., Makela et al., 2000; Nicolls et al., 2004). Figure 1-40 shows an example of the Arecibo ISR observations of large-scale TIDs with a period of 2 to 3 hours (Nicolls et al., 2004) that were generated by a series of geomagnetic substorms and propagated equatorward at very high velocities. At least three oscillations of the F layer were observed over Arecibo (18.3°N) with the final one driving the F peak to over 450 km and leading to small-scale structures indicative of plasma instabilities at dawn. In the lower panel of Figure 1-40 (Nicolls et al., 2004; Figure 1), the F-region vertical ion velocities along with polynomial fits are plotted, showing evidence for AGW downward phase progression with a velocity of ~200–300 m/s (Nicolls et al., 2004).
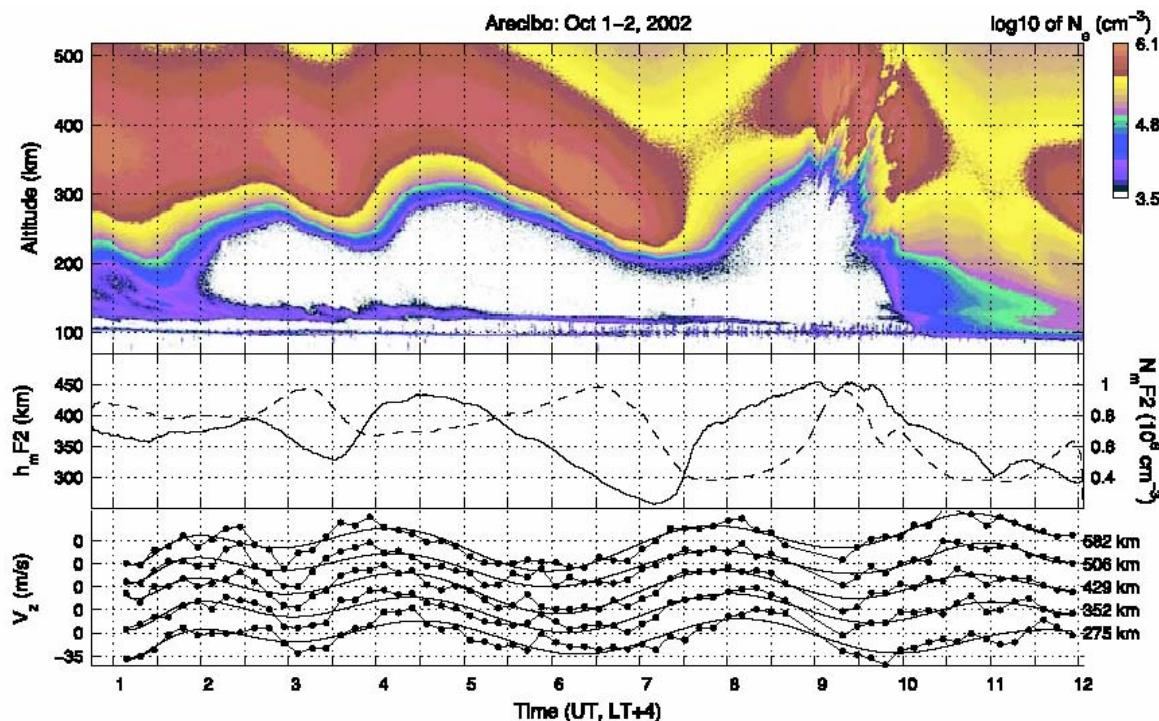
Figure 1-40: Summary of Arecibo ISR Data Showing Large-Scale TIDs for the Night of October 1–2, 2002. The top panel is $\log_{10}$ of electron density, the middle panel is h_mF_2 (solid) and N_mF_2 (dashed), smoothed over 12 minutes, and the bottom panel is vertical ion velocity (points) with polynomial fits (solid). The tick marks in the V_z panel are separated by 35 m/s (Nicolls et al., 2004).

e) Ionospheric Tomography

Radio tomography using satellite beacons was originally suggested by Austen et al. (1986). Several inversion techniques have been developed (e.g., Raymund et al., 1990, 1993 and 1994) and compared (Raymund, 1995). Further improvements of the computerized ionospheric tomography using GPS and low earth orbit (LEO) satellites have been achieved (Bernhardt et al., 1998). Satisfactory comparisons between the HF radar phase front observations and electron density measurements by EISCAT and tomographic techniques have been made (Pryse et al., 1995; Nygrén et al., 1996, Mitchell et al., 1998). Sequences of ionospheric images of TIDs can be analyzed to deduce the AGW spatial and temporal structure and the periods of the gravity waves.

f) Total Electron Content (TEC) from GPS Receiver Networks

While the ISR and tomographic techniques allow mapping the 2D structure of TIDs (Hunsucker, 1982; Hocke and Schlegel, 1996), additional techniques are needed to observe the 3D structure of TIDs. Total electron content (TEC) measurements in conjunction with other instrumentation like the MU radar (Shiokawa et al., 2002, 2003; Tsugawa et al., 2003) and ISRs (Nicolls et al., 2004) have been used to study large-scale TIDs. Total electron content (TEC) maps that respond to height-integrated TID effects are now provided by a dense GPS network (e.g., Coster et al., 2003). Individual receiver and satellite biases (Sardon et al., 1994; Wilson et al., 1999) are removed and slant TEC is converted to vertical TEC using a simple mapping function for an ionosphere stratified into thin layers. Difficulties of combining GPS TEC measurements with ISR observations of electron densities as a function of time and altitude are discussed by Nicolls et al. (2004), but useful techniques to study TEC perturbations due to TIDs have been developed (Saito et al., 1998; Ho et al. 1999; Afraimovich et al., 2000; Shiokawa et al., 2002a&b; 2003; Tsugawa et al. 2003; Nicolls et al., 2004). Recently,

TEC maps have been used to support observations of TIDs and tides by a super-resolution HF direction finding system (Hawlitschka, 2006).

As discussed by Nicolls et al. (2004), the TEC perturbation (TECP) at mid-latitudes caused by the ionospheric fluctuations associated with the neutral wind due to large-scale AGWs is not expected to be large. Using the GPS mapping technique, Nicolls et al. (2004; their Figure 3) showed nighttime TECPs of ~1-2 TECU (1 TECU = 10^{16} el/m^2) due to the large-scale TIDs on October 2, 2002 (Figure 1-41).
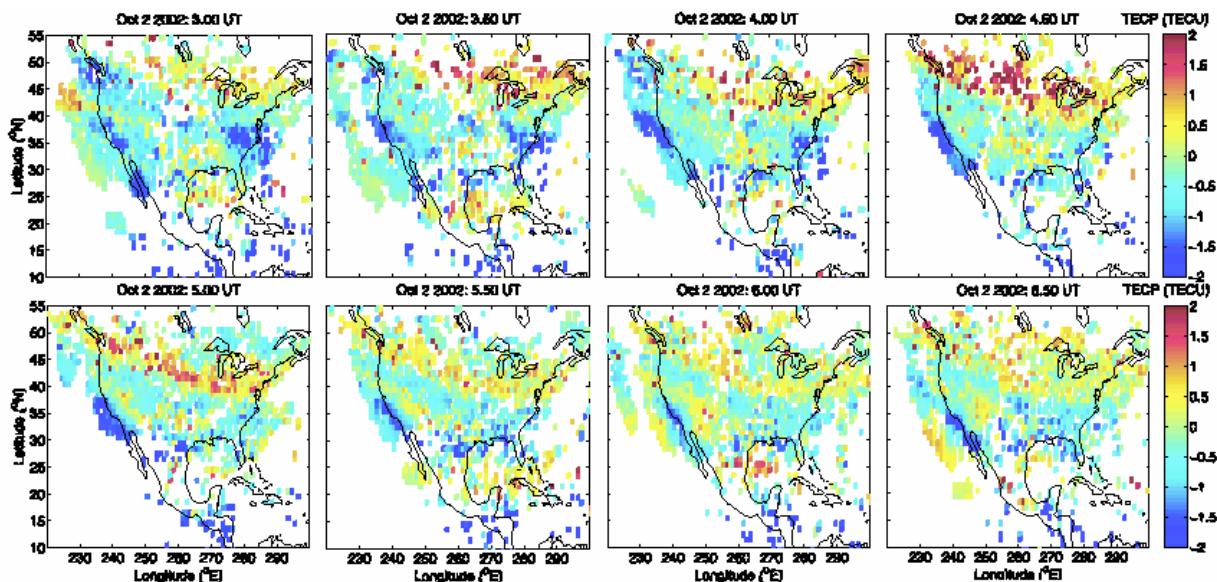
Figure 1-41: Examples of TECP Images Produced by the GPS Mapping Technique (Nicolls et al., 2004).

g) Airglow

Several naturally occurring airglow emissions originate in the mesosphere and lower thermosphere region 80-110 km (Hecht, 2004) as well as the bottom-side ionosphere region 100-300 km (Kubota et al., 2001). As the gravity waves propagate through the atmosphere they cause density and temperature variations, which are then detected by airglow imagers as quasi-periodic variations in the airglow emission intensity. Large- and medium-scale TIDs are commonly observed in the 630-nm airglow images as bands moving southwestward in the northern hemisphere (Mendillo et al., 1997; Shiokawa et al., 2000a) and have been detected at geomagnetically conjugate stations in the southern hemisphere (Otsuka et al., 2004; Shiokawa et al., 2005). Coordinated experiments using the Arecibo ISR, an airglow imager, and digital ionosondes addressed the electrodynamics of gravity waves in the thermosphere (Miller et al., 1997) and a new theory of gravity-wave-induced electric fields has been proposed (Miller, 1997; Kelley and Miller, 1997). However, at least one class of low-velocity TIDs is not well understood theoretically (Garcia et al., 2000).

1.3.8.4   Observations and Modeling of AGWs/TIDs and Their Sources

A causal relationship between the AGWs generated in the auroral zone (by Joule heating, Lorentz forcing or particle precipitation) and TIDs has been well established (Chimonas and Hines, 1970; Chimonas, 1970; Testud, 1970; Richmond, 1978; Hocke and Schlegel, 1996). The Worldwide Atmospheric Gravity-wave Studies (WAGS) program (see, e.g., Crowley and Williams, 1987; Williams et al., 1993) clearly demonstrated that large-scale TIDs observed at mid-latitudes originate in the auroral oval. The observed characteristic periods of TIDs (typically 30-70 min) are well correlated with auroral variations (electric

field and Joule heating) of similar periods. On the other hand, auroral fluctuations with time scales of 20 min or less are strongly attenuated in the observed TIDs. Thus the auroral plasma convection, which commonly exhibits quasi-periodic, recurrent bursts (e.g., Williams et al., 1992), is a likely source of gravity waves. Such a source then determines not only the magnitude but also the period of the gravity waves (Crowley and Williams, 1987).

Direct and earth-reflected gravity waves were theoretically described and distinguished by Francis (1974), who pointed out that the latter appear in the F region as wave packets (a number of nearly monochromatic wave cycles) while the former induce isolated (nonperiodic) TIDs, consistent with the modeling results by Millward et al. (1993a, b). Francis (1974) also suggested a dominance of upper-atmospheric (as opposed to tropospheric) sources of gravity waves. While the large-scale TIDs have been attributed to auroral electrojet surges, the identification of sources of medium-scale TIDs has remained difficult. Both auroral activity (e.g., Hunsucker, 1982, Lewis et al., 1996, Oyama et al., 2001) and tropospheric weather disturbances (e.g., Bertin et al., 1975; Waldock and Jones, 1987; Oliver et al., 1997) can generate medium-scale AGWs, particularly for TIDs that are observed at middle latitudes.

At high latitudes, Samson et al. (1990) suggested that the AGWs observed by SuperDARN originate near the ionospheric convection reversal boundary. Thus it has been hypothesized that the IMF variations that lead to auroral electrojet fluctuations generate AGWs/TIDs, but the actual identification of the distant sources of the observed TIDs has been posing difficulties in spite of the use of various observation techniques (Lewis et al., 1996; MacDougall et al., 1997; Hall et al., 1999, Oyama et al., 2001). One of the advantages of SuperDARN for the study of gravity waves is that it allows simultaneous observations of TIDs and of their high-latitude source activity. Various HF propagation modes frequently result in observations of the TID modulated ground scatter along with the ionospheric echoes at ranges immediately poleward of the ground scatter (Huang et al., 1998a; Sofko and Huang, 2000). A causality sequence of relationship among the IMF southward turnings, auroral electrojet currents, and gravity waves was reported for a case study (Huang et al., 1998b). The proximity of the observed TID signatures to the source region (auroral electrojet) indicated direct rather than ducted or earth-reflected gravity waves. Sofko and Huang (2000) associated each of the 100-min pulses observed in the ionospheric flow (electric field), the IMF and ground magnetic field with a pair of gravity waves (two wave pulses).

Modelling of gravity waves generated by enhancements in the ionospheric electric field shows that each electric field enhancement causes a Joule heating pulse, which in turn launches a gravity wave propagating equatorward and poleward of the source region (Millward et al., 1993a, b). This is consistent with some of the HF radar TID observations of TIDs which led Prikryl et al. (2005) to suggest that solar wind Alfvén wave coupling to the magnetosphere-ionosphere system is a source of auroral AGWs. Large scale traveling ionospheric/atmospheric disturbances generated by auroral surges have been observed and modeled (Hajkowicz, 1991; Balthazor and Moffett, 1997; Lee et al., 2004; Fujiwara and Miyoshi, 2006) to propagate as far as the equator. Medium-scale gravity waves can be ducted between the earth's surface and the lower thermosphere, propagate through the dissipationless lower atmosphere over long distances (Hocke and Schlegel, 1996) and leak from the lower atmosphere back to the thermosphere (Mayr et al., 1984a,b). Djuth et al. (2004) suggested that, near Arecibo, there was a continuum of gravity waves in the thermosphere, and that these showed "sets" of waves separated typically by 20-60 min. Livneh et al. (2007) used the incoherent scatter radar at the Arecibo Observatory to observe quasi-coherent continuous waves with periods of ~1 hour in the ionosphere. These are also typical periods of auroral AGWs (Bristow and Greenwald, 1997), which are likely generated by pulsed ionospheric convection (Prikryl et al., 2005) virtually at all times and sometimes observed as TIDs by SuperDARN (Bristow et al., 1996; He et al., 2004).

SuperDARN work on TIDs pursued studies of the relationship between auroral electrojet currents and TIDs (Sofko and Huang, 2000; Huang et al., 1998a,b) but also addressed the question of how one can deduce TID properties from SuperDARN ground scatter (Hall et al., 1999; Stocker et al., 2000;

MacDougall et al., 2001). In particular, the last named authors showed that the assumption that a TID must be located at approximately half the range of the radar ground scatter intensification that arises due to focusing by TIDs, may be valid only for low amplitude TIDs. Such an assumption, used in earlier SuperDARN studies of TIDs, implies that the TID horizontal phase velocity would be about half of the observed time rate of change of the distance to the ground backscatter intensification. This assumption of the factor 0.5 for making TID horizontal phase velocity estimations was challenged by Hall et al. (1999), who suggested that a more suitable value is 0.6. Further HF ray-tracing analysis and modeling of TIDs led MacDougall et al. (2001) to conclude that, depending on the TID amplitude, the appropriate proportionality factor may actually be close to 1.0, which almost doubles the horizontal TID phase velocities and wavelengths derived from the SuperDARN ground scatter signatures of TIDs.

Figure 1-42 shows an example of the SuperDARN Hankasalmi radar ground scatter power modulated by medium-scale gravity waves that were produced in the dayside auroral oval by ionospheric convection (Figure 1-39) that was pulsed by solar wind Alfvén wave coupling to dayside magnetopause (Prikryl et al., 2005). The slanted bands of strongly focused ground scatter power are due to equatorward-moving TID fronts in the F region. There is a one-to-one correspondence between the IMF-B_Y oscillations, the ground magnetic field perturbations and the TIDs.
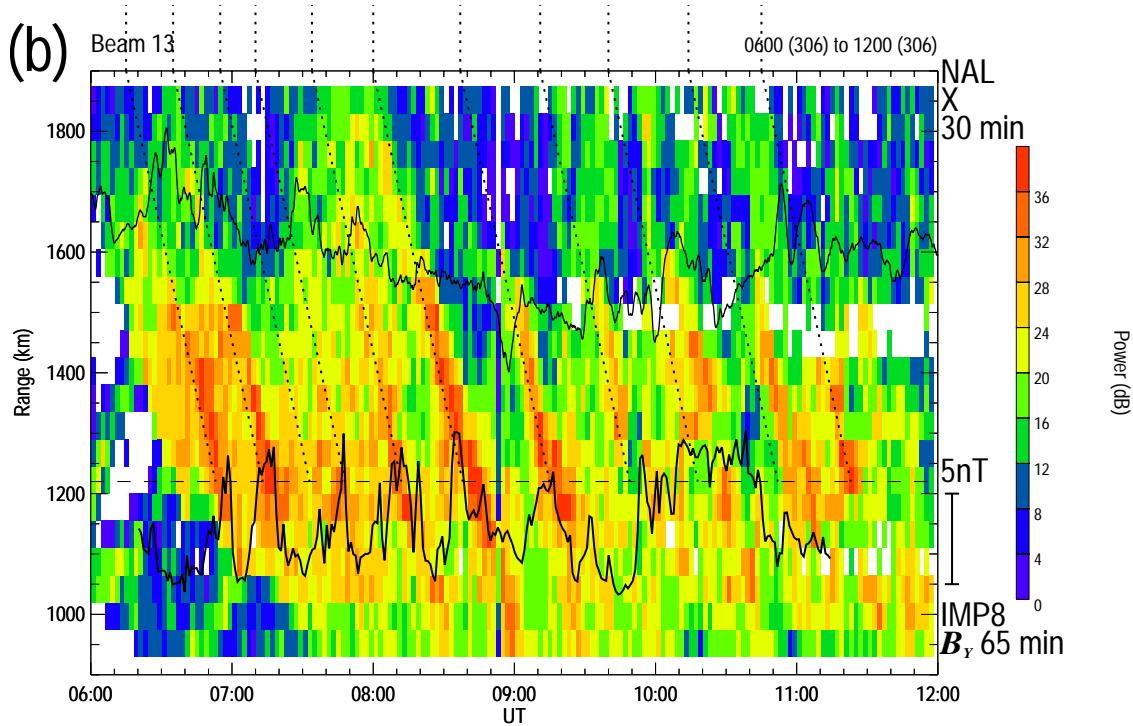
Figure 1-42: The SuperDARN Hankasalmi Radar Ground Scatter Power at Near Ranges.

Superposed are time series of the IMP-8 IMF BY (shifted by 65 min) and the Ny Åleslund (NAL) ground magnetic field X component (shifted by 30 min). The dotted lines along the medium-scale TID bands are superposed to show the correspondence between the ground magnetic field perturbations and the IMF-BY oscillations due to solar wind Alfvén waves (Prikryl et al., 2005).

While most of the above research has focused on identifying the wave source and wave properties, there have been studies, in addition to those mentioned in Sections 1.3.9.3 (d) and (e), that have attempted to determine the amplitude of the electron density fluctuation induced by the wave. By comparing the profiles obtained by ray-tracing calculations to model the received HF power with those measured by

SuperDARN, Bristow and Greenwald (1995) estimated that the peak density perturbations of TIDs were as great as 35%. This technique to estimate the electron density perturbations was extended by Stocker et al. (2000) who modeled the change in skip distance and obtained very similar amplitudes.

1.3.8.5   TID Effects on Transionospheric HF Propagation

While the reflection of HF radio waves from the bottom-side ionospheric layers that are modulated by TIDs are employed in TID detection techniques (Section 1.3.9.3), some of the HF power penetrates the ionosphere and can be detected by satellites. Transionospheric HF propagation can be significantly affected by TIDs. Transionospheric HF propagation experiments have been conducted in the past (James et al., 2006; James, 2006) and are planned for future satellite missions (Yau et al., 2006). James et al. (2006) investigated HF fades caused by multiple wave fronts detected by sounder receiver dipole antennas on board ISIS 1 and ISIS 2 spacecraft. Fades of the Faraday type, involving superposition of the ordinary and extraordinary wave components, as well as single-mode fades due to interference of multi-path signals, were observed.

Figure 1-43 shows rays at 9.303 MHz traced through a two-dimensional model ionospheric density based on tomographic data (Pryse et al., 1995) representing medium-scale TIDs with a horizontal wavelength of ~250 km. Figure 1-43 reveals the resulting shapes and orientations of mesoscale ionospheric irregularities, the most intense of which are elongated approximately parallel to the geomagnetic field. The low elevation rays are reflected and focused/defocused by the bottom-side ionospheric undulations caused by the TIDs. Because the density gradients of the elongated TID structure are approximately perpendicular to magnetic field, the poleward directed high-elevation rays (ray-gradient angles near 0º) suffer minimal refraction. In contrast, the ray-gradient angles are near 90º on the equatorward (left) side, so the structure is most effective in focusing or defocusing rays in that sector. Such rays were intercepted by the ISIS-2 spacecraft which observed interference fades as it was passing at altitude of 1400 km (James et al., 2006).
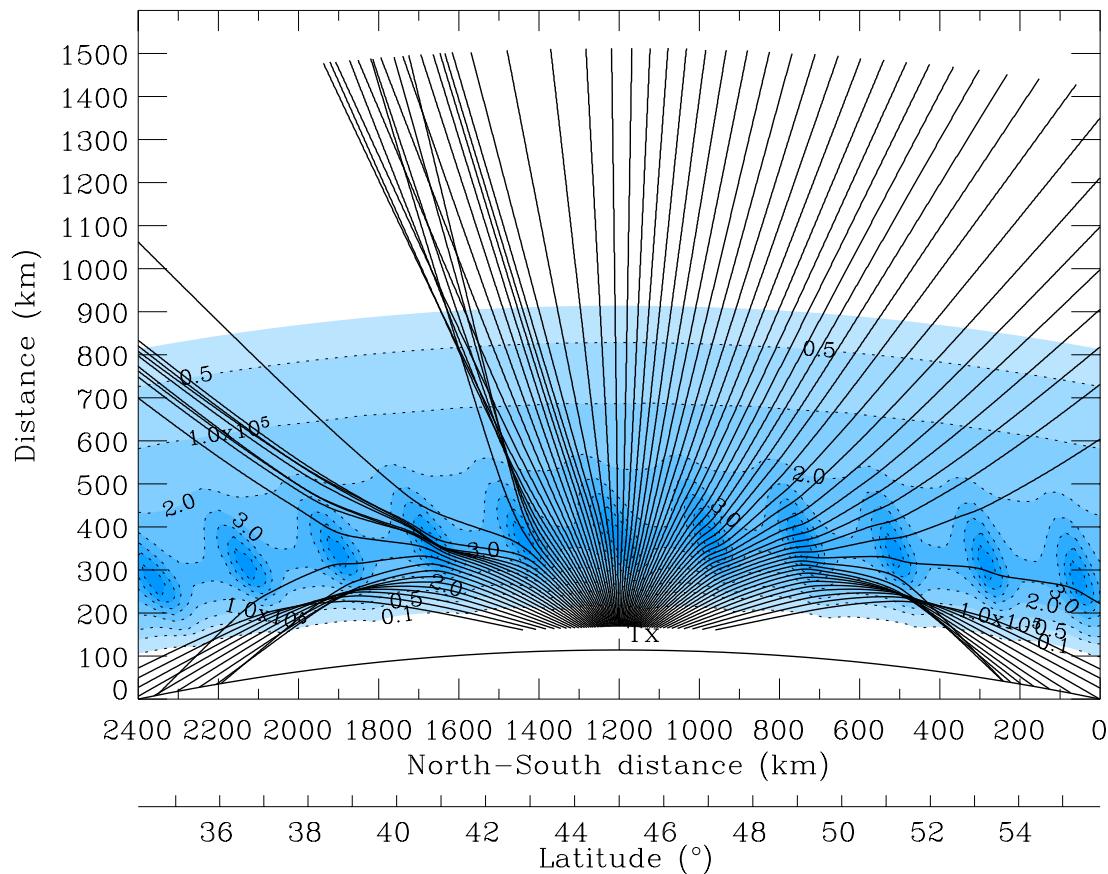
Figure 1-43: *O*-Mode Rays at 9.303MHz Traced in Two Dimensions Through a Model Ionosphere with Latitudinally Periodic Density Structure Resulting from Medium-Scale TIDs. The elevation of the lowest ray on the right side is 11° and increases in steps of 2°. The abscissa scale is distance along the chord of the Earth's surface, while the ordinate axis is at right angles to it. The latitude scale refers to the Earth's surface (James et al., 2006).

1.3.8.6   References

Afraimovich, E. L., Kosogorov, E. A., Leonovich, L. A., Palamartchouk, K. S., Perevalova, N. P., Pirog, O. M., Observation of large-scale traveling ionospheric disturbances of auroral origin by global GPS networks, *Earth, Planets and Space*, Volume 52, p. 669-674, 2000.

Austen, J. R., S. J. Franke, C.H. Liu, and K. C. Yeh, Application of computerized tomography techniques to ionospheric research, in *Radio beacon contribution to the study of ionization and dynamics of the ionosphere and corrections to geodesy*, Ed. A. Tauriainen, University of Oulu, Oulu Finland, Part 1, 25-35, 1986.

Balthazor, R. L. and R. J. Moffett, A study of atmospheric gravity waves and travelling ionospheric disturbances at equatorial latitudes, *Ann. Geophysicae*, Vol. 15(8), 1048-1056, 1997.

Bernhardt, P. A., McCoy, R. P., Dymond, K. F., Picone, J. M., Meier, R. R., Kamalabadi, F., Cotton, D. M., Chakrabarti, S., Cook, T. A., Vickers, J. S., Stephan, A. W., Kersely, L., Pryse, S. E., Walker, I. K., Mitchell, C. N., Straus, P. R., Na, H., Biswas, C., Bust, G. S., Kronschnabl, G. R., and Raymund, T. D.: Two-dimensional mapping of the plasma density in the upper atmosphere with computerized ionospheric tomography (CIT), *Phys. Plas.*, 5, 2010-2021, 1998.

Bertin, F., J. Testud, and L. Kersley, Medium-scale gravity waves in the ionospheric F-region and their possible origin in weather disturbances, *Planet. Space Sci.*, 23, 493-507, 1975.

Bristow, W. A., R. A. Greenwald, and J. C. Samson, Identification of high-latitude acoustic gravity wave sources using the Goose Bay HF radar, *J. Geophys. Res.*, 99, 319-331, 1994.

Bristow, W. A. and Greenwald, R. A.: Estimating gravity wave parameters from oblique high-frequency backscatter: modeling and analysis, *J. Geophys. Res.*, 100, 3639-3648, 1995.

Bristow, W. A. and Greenwald, R. A.: On the spectrum of thermospheric gravity waves observed by the Super Dual Auroral Radar Network, *J. Geophys. Res.,* 102(A6), 11585-11596, 1997.

Bristow, W. A., Greenwald, R. A., and Villain, J. P.: On the seasonal dependence of medium-scale atmospheric gravity waves in the upper atmosphere at high latitudes, *J. Geophys. Res.*, 101, 15685-15699, 1996.

Chimonas, G., and C. O. Hines, Atmospheric gravity waves launched by auroral currents, *Planet. Space Sci.*, 18, 565-582, 1970.

Chimonas, G., The equatorial electrojet as a source of long period travelling ionospheric disturbances, *Planet. Space Sci.*, 18, 583-589, 1970.

Chisham, G., Lester, M., Milan, S. E., Freeman, M. P., Bristow, W. A., Grocott, A., McWilliams, K. A., Ruohoniemi, J. M., Yeoman, T. K., Dyson, P. L., Greenwald, R. A., Kikuchi, T., Pinnock, M., Rash, J. P. S., Sato, N., Sofko, G. J., Villain J.-P., and Walker, A. D. M.: A decade of the Super Dual Auroral Radar Network (SuperDARN): scientific achievements, new techniques and future directions, *Surv. Geophys.*, 28, 33-109, DOI 10.1007/s10712-007-9017-8, 2007.

Coster, A. J., J. C. Foster, and P. J. Erickson, Space weather: Monitoring the ionosphere with GPS, *GPS World*, May, 40-45, 2003.

Crowley, G., and P.J.S. Williams, Observations of the source and propagation of atmospheric gravity waves, *Nature*, 328, 231-233, 1987.

Davies, K., Ionospheric Radio, Peter Peregrinus Ltd, London, 1990.

Djuth F. T., Sulzer, M. P., Gonzáles, S. A., Mathews, J. D., Elder, J. H., Walterscheid, R. L.: A continuum of gravity waves in the Arecibo thermosphere?, *Geophys. Res. Lett.*, 31, L16801, doi:10.1029/2003GL019376, 2004.

Francis, S. H., A theory of medium-scale traveling ionospheric disturbances, *J. Geophys. Res.*, 79, 5245-5260, 1974.

Fujiwara, H., and Y. Miyoshi, Characteristics of the large-scale traveling atmospheric disturbances during geomagnetically quiet and disturbed periods simulated by a whole atmosphere general circulation model, *Geophys. Res. Lett.*, 33, L20108, doi:10.1029/2006GL027103, 2006.

Garcia, F. J., M. C. Kelley, J. J. Makela, and C.-S. Huang, Airglow observations of mesoscale low-velocity traveling ionospheric disturbances at midlatitudes, *J. Geophys. Res.*, 105 (A8), 18,407-18,415, 2000.

Georges, T. M., HF Doppler studies of traveling ionospheric disturbances, *J. Atmos. Terr. Phys.*, 30, 735-746, 1968.

Grant, I. F., J. W. MacDougall, J. M. Ruohoniemi, W. A. Bristow, G. J. Sofko, J. A. Koehler, D. Danskin, and D. André, Comparison of plasma flow velocities determined by the ionosonde Doppler drift technique, SuperDARN radars, and patch motion, *Radio Sci.*, 30, 1537-1549, 1995.

Greenwald, R.A., K.B. Baker, J.R. Dudeney, M. Pinnock, T.B. Jones, E.C. Thomas, J.-C. Villain, J.-C. Cerrisier, C. Senior, C. Hanuise, R.D. Hunsucker, G. Sofko, J. Koehler, E. Nielsen, R. Pellinen, A.D.M. Walker, N. Sato, and H. Yamagishi, DARN/SUPERDARN A global view of the dynamics of high-latitude convection, *Space Sci. Rev.*, 71, 761-796, 1995.

Hajkowicz, L. A.: Auroral electrojet effect on the global occurrence pattern of large-scale traveling ionospheric disturbances, *Planet. Space Sci.*, 39, 1189-1196, 1991.

Hall, G.E., J.W. MacDougall, J.-F. Cecile, D.R. Moorcroft, and J.-P. St.-Maurice, Finding gravity wave source positions using the Super Dual Auroral Radar Network, *J. Geophys. Res.*, 104, 67-78, 1999.

Hargreaves, J. K., "The Solar Terrestrial Environment", Cambridge University Press, 1992.

Hawlitschka, S., Travelling ionospheric disturbances (TIDs) and tides observed by a super-resolution HF direction finding system, *Journal of Atmospheric and Solar-Terrestrial Physics*, 68(3-5), 568-577, 2006.

He, L.-S. , Dyson, P. L., Parkinson, M. L., and Wan, W.: Studies of medium scale travelling ionospheric disturbances using TIGER SuperDARN radar sea echo observations, *Ann. Geophysicae*, 22, 4077-4088, 2004.

Hecht, J. H., Instability layers and airglow imaging, *Rev. Geophys.*, 42, RG1001, doi:10.1029/2003RG000131, 2004.

Hines, C.O., Internal atmospheric gravity waves at ionospheric heights, *Can. J. Phys.*, 38, 1441-1481, 1960.

Ho, C. M., A. J. Mannucci, U. J. Lindqwister, X. Pi, and B. T. Tsurutani, Global ionosphere perturbations monitored by the worldwide GPS network, *Geophys. Res. Lett.*, 23(22), 3219-3222, 1996.

Hocke, K., and K. Schlegel, A review of atmospheric gravity waves and traveling ionospheric disturbances: 1982-1995, *Ann. Geophysicae*, 14, 917-940, 1996.

Huang, C. S., D. A. André, and G. J. Sofko, High-latitude ionospheric perturbations and gravity waves: 1. Observational results, *J. Geophys. Res.*, 103, 2131-2141, 1998a.

Huang, C. S., D. A. André, and G. J. Sofko, Observations of solar wind directly driven auroral electrojets and gravity waves, *J. Geophys. Res.*, 103, 23347-23356, 1998b.

Hunsucker, R. D., Atmospheric gravity waves generated in the high-latitude ionosphere: A review, *Reviews of Geophysics and Space Physics*, Vol. 20, 293-315, 1982.

Hunsucker, R.D., and L.H. Tveten, Large traveling-ionospheric-disturbances observed at midlatitudes utilizing the high resolution h.f. backscatter technique, *J. Atmos. Terr. Phys.*, 29, 909-916, 1967.

James, H. G., Effects on transionospheric HF propagation observed by ISIS at middle and auroral latitudes, *Advances in Space Research*, 38(11), 2303-2312, 2006.

James, H. G., R. G. Gillies, G. C. Hussey, and P. Prikryl, HF fades caused by multiple wave fronts detected by a dipole antenna in the ionosphere, *Radio Sci.*, 41, RS4018, doi:10.1029/2005RS003385, 2006.

Kelley, M. C., and C. A. Miller, Electrodynamics of midlatitude spread F 3. Electrohydrodynamic waves? A new look at the role of electric fields in thermospheric wave dynamics, *J. Geophys. Res.*, 102(A6), 11,539-11,548, 1997.

Kirchengast, G., K. Hocke, and K. Schlegel, Gravity waves determined by modeling of traveling ionospheric disturbances in incoherent-scatter radar measurements, *Radio Sci.*, 30(5), 1551-1567, 1995.

Kirchengast, G., Elucidation of the physics of the gravity wave-TID relationship with the aid of theoretical simulations, *J. Geophys. Res.*, 101, 13353-13368, 1996.

Kirchengast, G., K. Hocke, and K. Schlegel, The gravity wave-TID relationship: insight via theoretical model - EISCAT data comparison, *J. Atmos. Terr. Phys.*, 58, 233-243, 1996.

Kubota, M., H. Fukunishi, and S. Okano, Characteristics of medium- and large-scale TIDs over Japan derived from OI 630-nm nightglow observation, *Earth Planets Space*, 53, 741-751, 2001.

Lee, C.-C., J.-Y. Liu, M.-Q. Chen, S.-Y. Su, H.-C. Yeh, and K. Nozaki, Observation and model comparisons of the traveling atmospheric disturbances over the Western Pacific region during the 6-7 April 2000 magnetic storm, *J. Geophys. Res.*, 109, A09309, doi: 10.1029/2003JA010267, 2004.

Livneh, D. J., Seker, I., Djuth, F. T., Mathews, J. D.: Continuous quasiperiodic thermospheric waves over Arecibo, *Journal of Geophysical Research*, 112, A7, CiteID A07313, 2007.

Lewis, R.V., P.J.S. Williams, G.H. Millward and S. Quegan, The generation and propagation of atmospheric gravity waves from activity in the auroral electrojet, *J. Atmos. Terr. Phys.*, 58, 807-820, 1996.

MacDougall, J.W., G.E. Hall, and K. Hayashi, F region gravity waves in the central polar cap, *J. Geophys. Res.*, 102, 14513-14530, 1997.

MacDougall, J.W., D.A. Andre, G.J. Sofko, C.-S. Huang, and A.V. Koustov, Travelling ionospheric disturbance properties deduced from Super Dual Auroral Radar measurements, *Ann. Geophysicae*, 18, 1550-1559, 2001.

Makela, J. J., S.A. Gonzalez, B. MacPherson, X. Pi, M.C. Kelley, P.J. Sultan, Intercomparisons of total electron content measurements using the Arecibo incoherent scatter radar and GPS, *Geophys. Res. Lett.*, 27(18), 2841-2844, 2000.

Mayr, H. G., Harris, I., Varosi, F., and Herrero, F. A.: Global excitation of wave phenomena in a dissipative multiconstituent medium 1. Transfer function of the Earth's thermosphere, *J. Geophys. Res.*, 89, 10929-10959, 1984a.

Mayr, H. G., Harris, I., Varosi, F., and Herrero, F. A.: Global excitation of wave phenomena in a dissipative multiconstituent medium 2. Impulsive perturbations in the Earth's thermosphere, *J. Geophys. Res.*, 89, 10961-10986, 1984b.

Mendillo, M., J. Baumgardner, D. Nottingham, J. Aarons, B. Reinisch, J. Scali, and M. Kelley, Investigations of thermospheric-ionospheric dynamics with 6300 Å images from the Arecibo Observatory, *J. Geophys. Res., 102,* 7331, 1997.

Miller, C. A., W. E. Swartz, M. C. Kelley, M. Mendillo, D. Nottingham, J. Scali, and B. Reinisch, Electrodynamics of midlatitude spread F 1. Observations of unstable, gravity wave-induced ionospheric electric fields at tropical latitudes, *J. Geophys. Res.*, 102(A6), 11,521-11,532, 1997.

Miller, C. A., Electrodynamics of midlatitude spread F 2. A new theory of gravity wave electric fields, *J. Geophys. Res.*, 102(A6), 11,533-11,538, 1997.

Millward, G.H., S. Quegan, R.J. Moffett, T.J. Fuller-Rowell, and D. Rees, A modeling study of the coupled ionospheric and thermospheric response to an enhanced high-latitude electric field event, *Planet. Space Sci.*, 41, 45-56, 1993a.

Millward, G.H., R.J. Moffett, S. Quegan, and T.J. Fuller-Rowell, Effect of an atmospheric gravity wave on the midlatitude ionospheric F layer, *J. Geophys. Res.*, 98, 19173-19179, 1993b.

Millward, G.H., A resonance effect in AGWs created by periodic recurrent bursts in the auroral electric field, *Ann. Geophysicae*, 12, 94-96, 1994.

Mitchell, C. N., I. K. Walker, S. E. Pryse, I. Kersley, I. W. McCrea, T. B. Jones, First complementary observations by ionospheric tomography, the EISCAT Svalbard radar and the CUTLASS HF radar, *Ann. Geophysicae*, 16, 1519-1522, 1998.

Muldrew, D.B., An ionospheric ray-tracing technique and its application to a problem in long-distance radio propagation, *IRE Trans.*, Vol. AP-7, No. 4, 393-396, 1959.

Nicolls, M. J., M. C. Kelley, A. J. Coster, S. A. González, and J. J. Makela, Imaging the structure of a large-scale TID using ISR and TEC data, *Geophys. Res. Lett.*, 31, L09812, doi: 10.1029/2004GL019797, 2004.

Nygrén, T., Markkanen, M., Lehtinen, M., Tereshchenko, E. D., Khudukon, B. Z., Evstafiev, O. V., Pollari, P., Comparison of F-region electron density observations by satellite radio tomography and incoherent scatter methods, *Annales Geophysicae*, vol. 14, Issue 12, pp.1422-1428, 1996.

Oliver, W. L., Y. Otsuka, M. Sato, T. Takami, and S. Fukao, A climatology of F region gravity waves propagation over the middle and upper atmosphere radar, *J. Geophys. Res.*, 102, 14449-14512, 1997.

Otsuka, Y., K. Shiokawa, T. Ogawa, and P. Wilkinson, Geomagnetic conjugate observations of medium-scale traveling ionospheric disturbances at midlatitude using all-sky airglow imagers, *Geophys. Res. Lett.*, 31, L15803, doi: 10.1029/2004GL020262, 2004.

Oyama, S., M. Ishii, Y. Murayama, H. Shinagawa, S.C. Buchert, R. Fujii, and W. Kofman, Generation of atmospheric gravity waves associated with auroral activity in the polar F region, *J. Geophys. Res.*, 106, 18543-18554, 2001.

Prikryl, P., Muldrew, D. B., Sofko, G. J., and Ruohoniemi, J. M.: Solar wind Alfvén waves: A source of pulsed ionospheric convection and atmospheric gravity waves, *Ann. Geophysicae*, 23, 401-417, 2005.

Pryse, S. E., C. N. Mitchell, J. A. T. Heaton and L. Kersley, Travelling ionospheric disturbances imaged by tomographic techniques, *Ann. Geophysicae*, **13**, 1325-1330, 1995.

Raymund, T. D., Comparisons of several ionospheric tomography algorithms, *Ann. Geophysicae*, 13, 1254-1262, 1995.

Raymund, T. D., J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker, Application of computerized tomography to the investigation of ionospheric structures, *Radio Sci.*, 25, 771-789, 1990.

Raymund, T. D., S. E. Pryse, L. Kersley, and J. A. T. Heaton, Tomographic reconstruction of ionospheric electron density with European incoherent scatter verification, *Radio Sci.*, 28, 811-817, 1993.

Raymund, T. D., S. J. Franke, and K. C. Yeh, Ionospheric tomography: its limitation and reconstruction methods, *J. Atmos. Terr. Phys.*, 56, 637-657, 1994.

Richmond, A.D., Gravity wave generation, propagation, and dissipation in the thermosphere, *J. Geophys. Res.*, 83, 4131-4145, 1978.

Ruohoniemi, J. M. and Baker, K. B., Large-scale imaging of high-latitude convection with Super Dual Auroral Radar Network HF radar observations, *J. Geophys. Res.*, 103, 20797-20811, 1998.

Saito, A., S. Fukao, and S. Miyazaki, High resolution mapping of TEC perturbations with the GSI GPS network over Japan, *Geophys. Res. Lett.*, 25(16), 3079-3082, 1998.

Samson, J. C., R. A. Greenwald, J. M. Ruohoniemi, and K. B. Baker, High-frequency radar observations of atmospheric gravity waves in the high latitude ionosphere, *Geophys. Res. Lett.*, 16, 875-878, 1989.

Samson, J. C., R. A. Greenwald, J. M. Ruohoniemi, A. Frey, and K. B. Baker, J., Goose Bay radar observations of earth-reflected atmospheric gravity waves in the high-latitude ionosphere, *J. Geophys. Res.*, 95, 7693-7709, 1990.

Sardon, E., A. Rius, and N. Zarraoa, Estimation of the transmitter and receiver differential biases and the ionospheric total electron content from Global Positioning System observations, *Radio Sci.*, 29(3), 577-586, 1994.

Shiokawa, K., Y. Otsuka, M. K. Ejiri, Y. Sahai, T. Kadota, C. Ihara, T. Ogawa, K. Igarashi, S. Miyazaki, and A. Saito, Imaging observations of the equatorward limit of midlatitude traveling ionospheric disturbances, *Earth Planets Space*, 54, 57-62, 2002a.

Shiokawa, K., Y. Otsuka, T. Ogawa, N. Balan, K. Igarashi, A. J. Ridley, D. J. Knipp, A. Saito, and K. Yumoto, A large-scale traveling ionospheric disturbance during the magnetic storm of 15 September 1999, *J. Geophys. Res.*, 107(A6), 1088, doi: 10.1029/2001JA000245, 2002b.

Shiokawa, K., et al., Thermospheric wind during a storm-time large-scale traveling ionospheric disturbance, *J. Geophys. Res.*, 108(A12), 1423, doi: 10.1029/2003JA010001, 2003.

Shiokawa, K., Y. Otsuka, T. Tsugawa, T. Ogawa, A. Saito, K. Ohshima, M. Kubota, T. Maruyama, T. Nakamura, M. Yamamoto, and P. Wilkinson, Geomagnetic conjugate observation of nighttime medium-scale and large-scale traveling ionospheric disturbances: FRONT3 campaign, *J. Geophys. Res.*, 110, A05303, doi:10.1029/2004JA010845, 2005.

Sofko, G. J. and C.S. Huang, SuperDARN observations of medium-scale gravity wave pairs generated by Joule heating in the auroral zone, *Geophys. Res. Lett.*, 27, 485-488, 2000.

Stocker, A.J., N.F. Arnold, and T.B. Jones, The synthesis of travelling ionospheric disturbances (TID) signatures in HF radar observations using ray tracing, *Ann. Geophysicae*, 18, 56-64, 2000.

Testud, J., Gravity waves generated during magnetic substorms, *J. Atmos. Terr. Phys.*, 32, 1793-1805, 1970.

Tsugawa, T., A. Saito, Y. Otsuka, and M. Yamamoto, Damping of large-scale traveling ionospheric disturbances detected with GPS networks during the geomagnetic storm, *J. Geophys. Res.*, 108(A3), 1127, doi: 10.1029/2003JA09433, 2003.

Tsugawa, T., A. Saito, and Y. Otsuka, A statistical study of large-scale traveling ionospheric disturbances using the GPS network in Japan, *J. Geophys. Res.*, 109(A6), A06302, doi: 10.1029/2003JA010302, 2004.

Tveten, L.H., Ionospheric motions observed with high-frequency backscatter sounders, J. of Research of the National Bureau of Standards, *D. Radio Propagation,* 65D, 115-127, 1961.

Waldock, J. A., and T.B. Jones, Source regions of medium scale traveling ionospheric disturbances observed at mid-latitudes, *J. Atmos. Terr. Phys.*, 49, 105-114, 1987.

Williams, P.J.S., R.V. Lewis, T.S. Virdi, M. Lester, E. Nielsen, Plasma flow bursts in the auroral electrojets, *Ann. Geophysicae*, 10, 835-848, 1992.

Williams, P.J.S., T.S. Virdi, R.V. Lewis, M. Lester, A.S. Rodger, I.W. McCrea, and K.S.C. Freeman, Worldwide atmospheric gravity-wave study in the European sector 1985-1990, *J. Atmos. Terr. Phys.*, 55, 683-696, 1993.

Wilson, B. D., C. H. Yinger, W. A. Feess, and C. Shank, New and improved - the broadcast inter-frequency biases, *GPS World*, 10, 56-66, 1999.

Yau, A. W., James, H. G., Liu, W., The Canadian Enhanced Polar Outflow Probe (e-POP) mission in ILWS, *Advances in Space Research*, Volume 38, Issue 8, p. 1870-1877, 2006.

Yeh, K. C. and Liu, C. H.: Acoustic-gravity waves in the upper atmosphere, *Rev. Geophys. Space Phys.*, 12, 193-216, 1974.

Chapter 2 – INSTRUMENTATION

2.1   THE IONOSONDE

(John MacDougall, Dept. of Electrical and Computer Engineering, University of Western Ontario, London, Ont. Canada N6A 5B9)

2.1.1   Ionospheric Sounding

An Ionosonde is an HF radar used for obtaining electron density profiles of the ionosphere. It basically works on the principle that a radio wave will be totally reflected if the radio frequency is equal to the characteristic frequency of the ionosphere which depends on plasma frequency, $f_p = (80.5N)^{1/2}$ Hz, where N is the electron density in electrons per m^3. Thus knowing the height of reflection and the radio frequency one has a measurement of the electron density at that height.

This simplified principle is somewhat more complex in practice since the group velocity of the radio wave is slower in the ionosphere so the height obtained using the time delay of the echo and assuming that the radio wave travels at the speed of light is an overestimation of the actual height of reflection. Also, the ionosphere is birefringent for radio waves so that the reflection conditions for one mode, the extraordinary wave, are different from the reflection conditions for the other mode, the ordinary wave. The electric field of the ordinary wave is left hand circularly polarized when viewed along the direction of the terrestrial magnetic field. For ordinary waves the reflection condition is simple in that the wave is reflected at the level where the frequency just equals the plasma frequency as defined above. It is therefore necessary to have a way of separating out the ordinary wave reflections when calculating ionospheric densities from the echoes obtained by an ionosonde. Having separated out the O mode echoes an estimation of the electron density profile can be obtained using a "true height analysis" program. The best known true height analysis program is Polan [Titheridge, 1985, 1998].

2.1.2   Ionosonde Specifications

From these, and other, considerations one can list the specifications for an ionosonde:

1) **Frequency Range:** Based on maximum ionospheric densities up to a few times 1.0×10^{12} electrons per m^3 => 9 MHz plasma frequency. Thus ionosondes that will be used for obtaining overhead electron densities typically should have an upper frequency ~20 MHz. The minimum frequency is generally determined by considerations such as antenna size, licensing restrictions, and ionospheric absorption effects (higher at low frequencies). Minimum ionosonde frequencies of the order of 1 MHz are typical.

2) **Range:** Since an ionosonde can only obtain echoes where its frequency matches the plasma frequency it is limited to observations of the 'bottomside' of the ionosphere. (When it uses a frequency higher than the maximum plasma frequency of the ionosphere the transmitted pulse passes through the ionosphere without reflection.) The height range of interest is therefore from about 90 kilometers up to the order of 1000 km.

3) **Range Resolution:** If the ionosonde is just being used for electron density profiles then range resolution of a few kilometers is usually adequate since it is only when there is Sporadic E present that there is major ionospheric structure with a scale size smaller than a few kilometers. Note that high range resolution implies shorter pulses and therefore wider receiver bandwidth and hence higher receiver noise levels. For this reason most ionosondes are not designed to have range resolution better than a few kilometers.

4) **Polarization Separation:** Some way of separating O and X circular polarization is necessary. This can be done by using antenna arrangements, and electronics, that only transmit the O or X mode. An alternative method is to combine signals from perpendicularly orientated receiving antennas in such a way that the two modes of the echoes can be determined. Of course visually one can usually distinguish between the O and X modes because the X mode is reflected with about ½ electron gyrofrequency (typically ~1/2*1.4 MHz) higher frequency than the O mode.

5) **Transmitter Power:** Older model ionosondes used large peak transmitter power of the order of 10 kW [Hunsucker, 1991]. This was necessary in order to get acceptable signal-to-noise performance. Newer ionosondes generally use some form of signal coding/processing to improve the signal-to-noise.

The following is a basic calculation of the signal-to-noise pertinent to an ionosonde. At HF the noise levels on the ground are very high. There are also other users on these frequencies which make the noise situation even worse. For this calculation an average noise temperature of 10^8 K will be used. Assuming a 30 kHz receiver bandwidth (typical ionosonde value) the received noise power would be 4×10^{-11} W. If a signal-to-noise ratio of 20 dB is required then the received power should be 4×10^{-9} W. The received power for a transmitted power P_T is $P_R = P_T\, G_T\, G_R\, \lambda^2\, /(4\pi R)^2$ where G_T is the gain of the transmitter antenna, G_R is the gain of the receiver antenna, λ is the wavelength, R is the range which for specular (total) reflection is twice the actual range. (If there is ionospheric absorption of the signal the power received would be lower.) For calculation use: $G_T = G_R = 1$, $\lambda = 100$ m, $R = 2\times300$ km and from these obtain a required $P_T = 23$ W. Thus, theoretically, an ionosonde requires only very low peak transmitter power even without coding/signal processing.

In general, ionosondes use very much higher peak power than 23 W. There are a number of reasons for this. The most obvious reason is that the gains of the transmitting and receiving antennas are usually much less than 1. This is because it is expensive to build wideband HF antennas that have an upward pointing beam and have good gain. Assuming that the antenna gains are only 1/10 then the peak power required becomes 2.3 kW which is comparable with the 10 kW power used for older ionosondes.

Because it is easy to obtain power levels of a few hundred watts using solid state amplifiers, and it is convenient and inexpensive to do digital coding/signal processing, most modern ionosondes have peak power of the order of 100-1000 W and employ a great deal of coding/signal processing techniques. As a typical example, using 13 bit Barker code the signal-to-noise is improved by $10\times\log(13) = 11$ dB, and employing coherent (FFT) averaging on 8 samples gives another 9 dB so a total improvement of 20 dB is easily obtained. One could obtain further improvement by using a longer code [Huang and MacDougall, 2005] but the requirement to obtain echoes from a height of ~90 km means that the total length of the coded pulse cannot be longer than ~600 μs unless the transmitter and receiver are located sufficiently far apart that the transmitted pulse will not saturate the receiver.

Further signal-to-noise improvement can be obtained by coherently averaging a number of samples. The signal-to-noise improvement is 10*log(number of samples). However there is a problem in this if there is an appreciable Doppler shift in the received echoes since, for a long series of samples, the phase shift of the later samples, relative to the first samples, could become sufficiently large that when one coherently adds them they might subtract rather than add. It is of course possible to overcome this problem by progressively phase shifting the samples to remove the Doppler phase shifts. The problem with doing this is that one does not know a value for the Doppler shift. However, one can get around this problem by doing an FFT instead of coherent averaging. An FFT is effectively the same, if one chooses the correct frequency bin, to removing the Doppler phase shifts by progressively phase shifting the samples before coherently averaging them. This means that at least theoretically, one can use as large a number of samples for the FFT as desired for signal-to-noise improvement. Usually the number of samples is curtailed by the desire to minimize the time needed to obtain the echo results for each ionosonde frequency. A good discussion of some of these signal

processing techniques as used in ionosondes is the Digisonde web site [http://ulcar.uml.edu/digisonde_dps. html]. See also Reinisch [1996]. There is an abundance of additional information about ionosondes on the Internet.

2.1.3   Ionogram Samples

The behaviour, and hence ionogram appearance, of the earth's ionosphere are different for different locations. A simplified division of the regions showing different ionogram behaviour are: mid-latitudes, auroral zone, polar cap region, equatorial region. A few examples from these regions are shown here.

Figure 2-1 shows samples of ionograms from Delaware (43° N, 81° W) for 2006 September 8. Times marked are UT: subtract 5 hours for local time. Daytime ionograms are shown in the two left hand panels, and nighttime in the two right hand panels. In daytime sunlight directly creates the F_1 layer. Above the F_1 layer one sees the F_2 layer which essentially is due to the reduction of recombination at greater heights. There is also a weak E region, also sunlight produced. During daytime echoes, particularly on lower frequencies, may be weak due to radio wave absorption in the lower D region. After sunset the F_1 layer disappears and a progressively weaker F_2 layer remains throughout the night, as seen in the two nighttime panels.
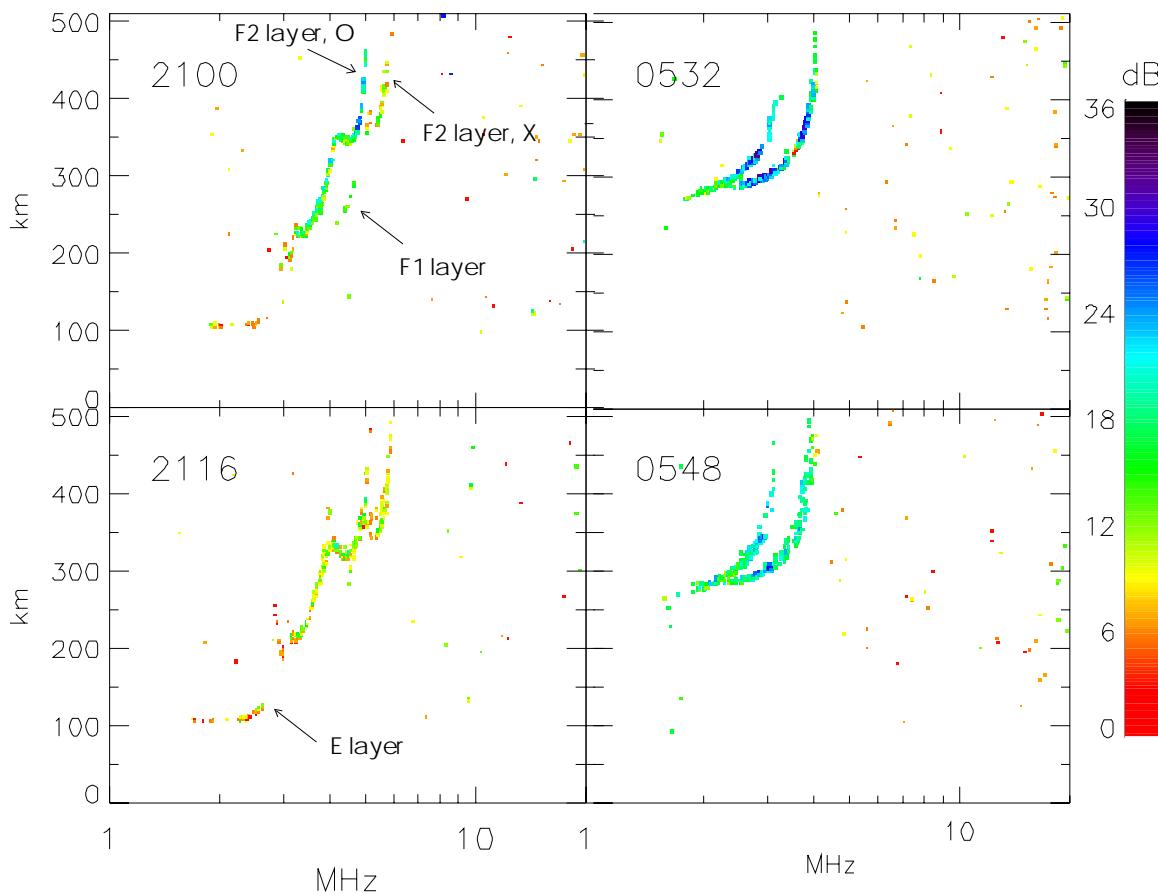
Figure 2-1: Mid-Latitude Ionograms.

Figure 2-2 shows ionograms from Rabbit Lake (58^oN, 104^oW) for early nighttime 1995 Nov. 8. For these ionograms one sees E region (heights 90 – 150 km) echoes from a layer that is created by energetic precipitation. The appearance of the layer varies notably between the samples. Sometimes one sees a thick

layer (as at 0200) with the O and X traces being distinguishable, whereas at other times one just sees an irregular patch. If the precipitation becomes very energetic it creates a D region below 90 km and this absorbs the radio signals so the ionogram may not show any sign of echoes.
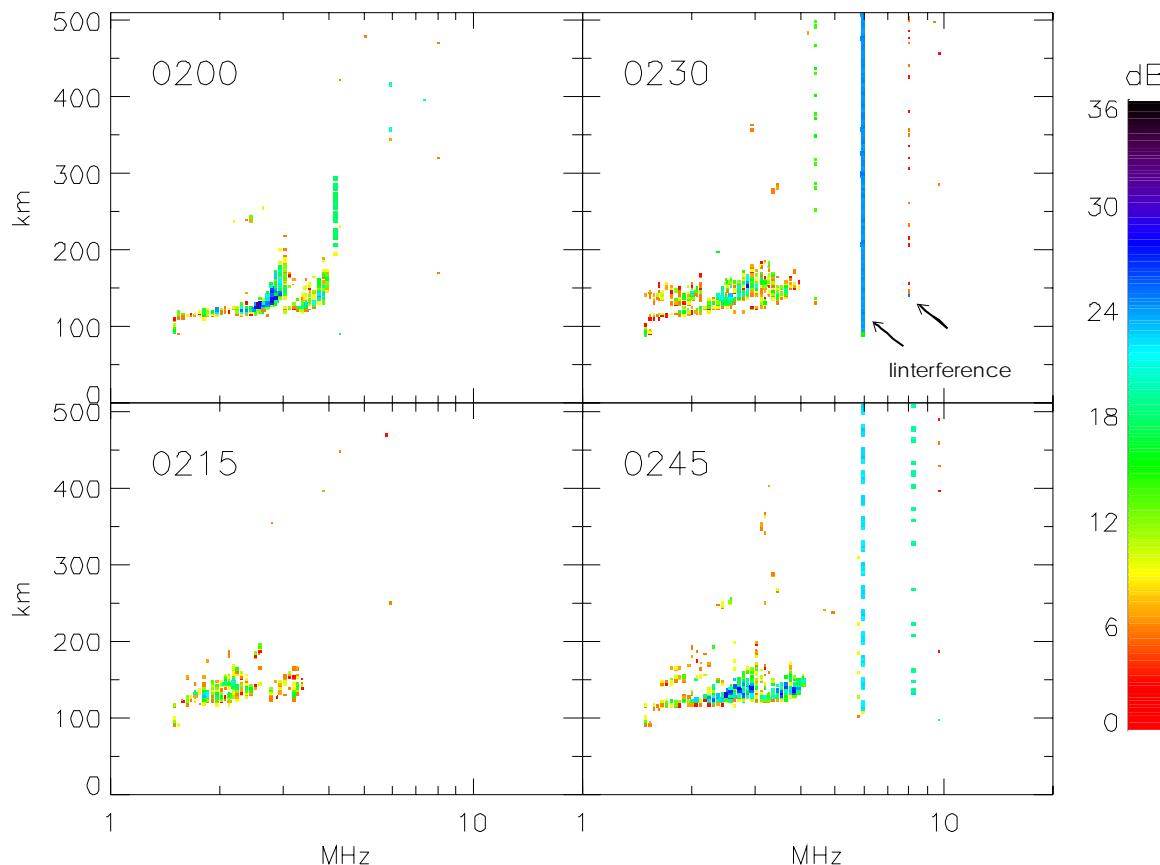
Figure 2-2: Auroral Zone Ionograms.

Figure 2-3 shows ionograms from Resolute Bay, 2007 January 8. Times in UT are marked on the ionograms. Resolute Bay (75° N, 95° W) is in the polar cap region. Note the spreading which is typical for polar cap ionograms. On only one of the 4 ionograms can one clearly distinguish the Ordinary (O) and Extraordinary (X) traces. Other features such as Sporadic-E (Es), vertical interference streaks, off vertical (non overhead) echoes, and 2nd hop echoes (reflected from ionosphere and then from ground and back to 2nd ionospheric reflection) are marked.
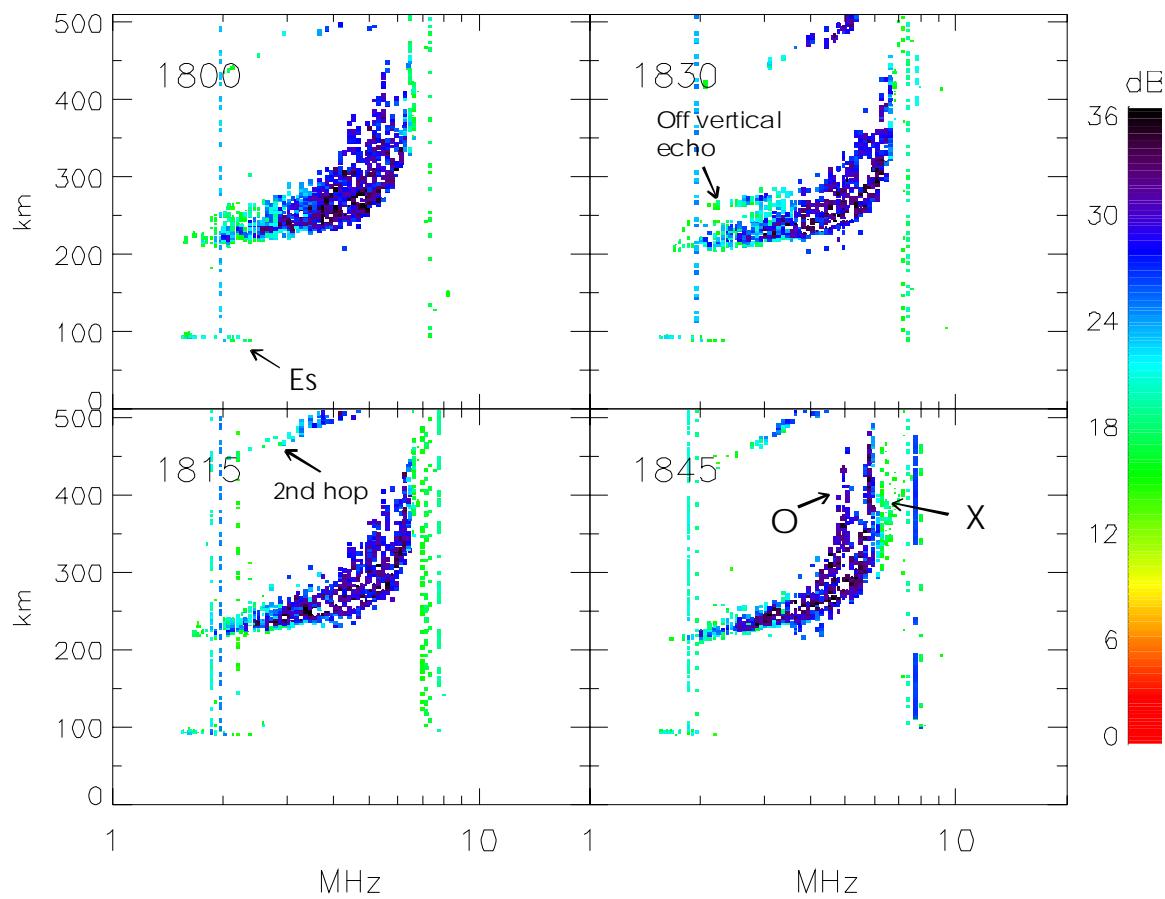
Figure 2-3: Polar Cap Ionograms.

Figure 2-4 shows ionograms from Fortaleza (4°S, 38°W) during daytime 1996 Sept. 13. The F_1 and F_2 layers can be clearly seen (compare with Figure 2-1). The Q-type Sporadic-E marked is a type that is only seen very close to the magnetic dip equator. Another difference between equatorial locations and mid-latitudes is that at the equator the F region is often pushed up to great heights (this sample does not show this) so the height scale needs to go to ~1000 km.
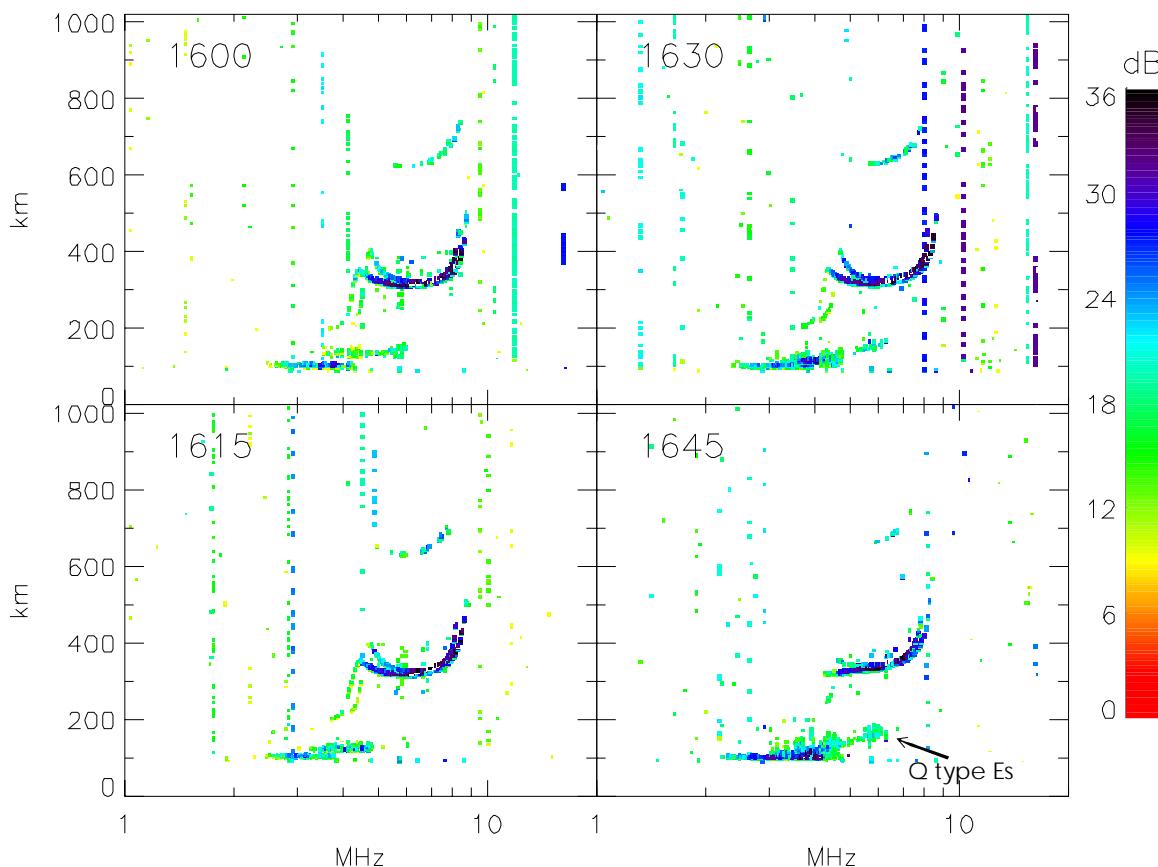
Figure 2-4: Equatorial Ionograms.

2.1.4   Other Ionosonde Measurements

Modern ionosondes usually do a variety of measurements other than just recording ionograms. Some of the additional measurements that are commonly done are discussed here:

*   **Oblique Ionograms:** This measurement is used to confirm/determine propagation conditions between two locations. It requires synchronized ionosonde receivers and transmitters at two locations, synchronization now being relatively easy because of GPS. The measurement usually requires the ionosondes used at the two ends of the path to operate on frequencies higher than those used for vertical ionospheric measurements, and requires measurements for longer ranges.

*   **Ionospheric Velocities (HF Doppler Technique):** By taking a series of samples and from the series determining the Doppler shifts one can determine the ionospheric velocities. This measurement only works well for giving the ionospheric vector velocity if there are many ionospheric reflecting 'glints'. If there is Spread-F there are normally many glints and therefore the possibility of good ionospheric vector velocities using this technique.

*   **Ionospheric Absorption:** Ionosonde data are not commonly analysed for absorption, but usually the echo strengths are recorded so the temporal variation can easily be obtained. Since this method can measure absorption at many HF frequencies it is more sensitive than the usual riometer measurements when absorption is weak.

*   **Traveling Ionospheric Disturbances:** TIDs are launched by various atmospheric/ionospheric disturbances, usually atmospheric gravity waves. They usually show up clearly in the F region where they cause variations in ionospheric height and density. To obtain information about the

TID propagation requires the combination of measurements from several separated ionosondes which need to operate at a rate sufficient to measure the TID variations which typically have periods ranging from a few minutes in the E region to tens of minutes in the F region.

Future ionosondes will be improved relative to current models by, in particular, the use of 'software radio' technology which will make it much simpler to build a high quality receiver for HF. This is especially true since at HF the incoming signal, after slight amplification, can be directly digitized. This software radio technology eliminates the 'image' problem in HF receivers. The image problem arises as follows: An ionosonde receiver is required to receive over a range of frequencies, 1 – 20 MHz being typical. In a standard superheterodyne receiver the incoming radio frequency signal is mixed (multiplied) with the local oscillator signal and the frequency difference between the incoming signal and local oscillator is the intermediate frequency, $f_{IF} = f_{RF} - f_{LO}$ or $f_{RF} = f_{LO} + f_{IF}$. Most of the receiver amplification is done in the intermediate frequency amplifier. Now there is a problem in this because it is easily shown that the above equation should actually be $f_{RF} = f_{LO} \pm f_{IF}$ so the receiver can simultaneously receive 2 frequencies. One of these frequencies is the 'correct' RF frequency and the 2nd is the image. Because the receiver has to operate over a range of RF frequencies the image will be a problem unless the IF frequency is selected so that the image frequency is not in the 1-20 MHz band of frequencies. This requires the use of a high IF frequency (>20 MHz) and make for a somewhat convoluted arrangement of intermediate frequency stages in some receivers. In a software radio this problem does not occur because, effectively, one can choose an IF frequency = 0 so there is no image.

As a basic tool for ionospheric monitoring and measurements the ionosonde has been a fundamental instrument for more than 50 years. The modern ionosonde is much cheaper and runs with minimal maintenance as compared to the older ionosondes. The modern ionosonde can also do ancillary measurements, some of which are listed above.

2.1.5   References

Huang, J. and J. MacDougall, Legendre coding for digital ionosonde, Radio Sci. 40 (RS4013), doi: 10.1029/2004RS003123, 2005.

Hunsucker, R. D., Radio Techniques for Probing the Terrestrial Ionosphere, Physics and Chemistry in Space Vol. 22, Springer-Verlag, 1991.

Reinisch, B.W., "Modern Ionosondes", Modern Ionospheric Science, H. Kohl, R. Ruster, and K. Schlegel (eds.), pp. 440-458, 1996.

Titheridge, J. E., Ionogram Analysis with the Generalized Program POLAN, World Data Center A report UAG-93, 1985.

Titheridge, J. E. The real height analysis of ionograms: a generalized formulation. Radio Sci. 23, 831-849, 1998.

2.2   TOPSIDE SOUNDING

(Gordon James, Communications Research Centre)

2.2.1   Introduction

The concept of a topside "ionosonde", or similar, radar on an artificial satellite began to be developed in the late 1950s, not long after the Soviet launch of Sputnik 1. Canada and the U.S.A. agreed to work on fixed- and swept-frequency topside sounders in 1959. The first swept-frequency sounder spacecraft

Alouette-I was launched in 1962, followed by a fixed-frequency sounder on Explorer-20 in 1964. The success of swept-frequency sounding with Alouette-I led to extensive international participation in research on the resulting data. There followed the launching of three more Canada-U.S. sounders: Alouette II, ISIS I, and ISIS II.

The major findings of the Alouette-ISIS scientific program fall mostly into 3 major areas [IEEE, 1969; Jackson et al., 1980]:

1) A wealth of global information has been gathered about the electron density distribution from the F-region electron peak upwards, not obtainable in any other way. Techniques were developed for deconvolving topside sounder ionograms, using cold-plasma and geometric-optics concepts, into profiles of density with altitude that, when combined in continuous series, provided two-dimensional views of the structure of the topside ionosphere, over a range of scale sizes;

2) A variety of short wavelength electrostatic waves requiring hot-plasma concepts and sometimes the consideration of large-signal nonlinearity for their explanation have been documented; and

3) Understanding has been enlarged about the nature of a variety of spontaneous electromagnetic radiations emanating from free-energy sources in space or on the ground.

The mini-review of Jackson et al. [1986] on topside sounding up to the 1980s summarizes the techniques used to obtain topside ionograms and reduce them to electron profiles. The Jackson et al [1986] review is relevant to the present paper in its explanation of the most important differences between satellite-based ionograms and ground-based ionograms described in the previous section.

Sounders were placed on later spacecraft by other national agencies:

1) The ISS-b (Japan – 1978) [Wakai, 1980] sounder data were used to study global F-region distribution of density.

2) The Soviet Interkosmos-19 (1979) and Cosmos -1809 sounders (1986) [Shuiskaya et al., 1990] provided data for research on plasma nonlinearities triggered in the intense near fields of transmitting antennas.

Jackson's [1986] bibliography reported several hundred papers in the archival literature based on Alouette-ISIS data. The story, however, has not ended. The systematic digitization of Alouette and ISIS ionograms and the subsequent use of this accessible data attest to the ongoing relevance of sounder data for modern research subjects [Benson and Osherovich, 2004].

Low-power sounders, also called "relaxation" sounders, have been productive in extending our understanding of electrostatic resonance waves that reflect at short distances from an active spacecraft. The resulting resonance observations were processed into data sets of the spatial dependence of plasma parameters in various regions coverable by Earth-orbiting spacecraft. Examples of these are the French GEOS-1 and -2 [Etcheto et al., 1981], the Japanese EXOS series [Oya and Ono, 1981] and the ESA Cluster [Trotignon et al., 2004].

2.2.2   The Radio Plasma Imager on the IMAGE Spacecraft

In 2000, NASA launched the Radio Plasma Imager (RPI) on the Imager for Magnetopause-to-Aurora Global Exploration (IMAGE) spacecraft. The RPI has been employed for radio sounding from magnetospheric altitudes up to 8 earth radii, relaxation sounding, and passive plasma wave observations [Green and Reinisch, 2003]. A method for determining the density distribution of the plasma from either direct or field-aligned echoes has been applied to the determination of the evolving density structure of the polar cap and the plasmasphere, under a variety of geomagnetic conditions. The RPI plasma resonance observations provide data on the ambient plasma densities and on electrostatic wave resonances at large radial distances. The RPI

long antennas and noise receivers yielded new evidence about the nature of spontaneous continuum radiation and auroral kilometric radiation. Coordination of RPI and the Extreme Ultraviolet (EUV) imager on IMAGE resulted in discoveries about the relation of kilometric continuum to the shape of the plasmasphere. Finally, RPI has been used to successfully test the feasibility of magnetospheric tomography, using intersatellite transmission to the Wind spacecraft.

2.2.3 Future

It seems likely that active ionospheric RF sounding will continue to be pursued. The MARSIS radar on the MARS EXPRESS spacecraft was conceived primarily for surface sounding of that planet but turned up some ionospheric topside signatures that have been interpreted with notions used on the terrestrial sounder observations [Gurnett et al, 2006]. The Enhanced Polar Outflow Probe (ePOP) experiment [Yau et al., 2006] will be launched in 2008 on the Earth-orbiting CASSIOPE spacecraft. CASSIOPE will venture into a new version of transionospheric HF sounding, wherein an ionosonde on the ground will both provide conventional soundings of the bottomside and signals transmitted to CASSIOPE. The resulting observations will be combined for imaging of ionospheric density structure.

2.2.4 References

Benson, R.F. and V.A. Osherovich, Application of ionospheric topside-sounding results to magnetospheric physics and astrophysics, *Radio Science, 39*, RS1S28, doi:10.1029/2002RS002834, 2004.

Etcheto, J., H. de Feraudy, and J.H. Trotignon, Plasma resonance stimulation in space plasmas, *Adv. Space Res, 1*(2), 183-196, 1981.

Green, J.L. and B.W. Reinisch, *Space Science Reviews*, Special issue on IMAGE results, 109, 183-210, 2003.

Gurnett, D. , R. Huff , D. Kirchner, D. Morgan, A. Persoon, T. Averkamp, F. Duru, F. Akalin, E. Nielsen, A. Safaeinili, J. Plaut, and G. Picardi, Mars Express radar soundings of the ionosphere of Mars, COSPAR Scientific Assembly 2006, Beijing, 16-23 July, Paper C3.2-0007-06, 2006.

IEEE, Special issue on topside sounding and the ionosphere, *Proc. IEEE*, *57*, 6, 1969.

Jackson, J.E., E.R. Schmerling, and J.H. Whitteker, "Mini-review on topside sounding," *IEEE Trans. Ant. Propag., 28*(2*)*, 284-288, 1980.

Jackson, J.E., Alouette-ISIS program summary, National Space Science Data Center, Doc. no. NSSDC/WDC-A-R&S 86-09, Greenbelt, MD, U.S.A., 1986.

Oya, H. and T. Ono, Stimulation of plasma waves in the magnetosphere using satellite Jikiken /EXOS-B/, *Adv. Space Res., 1*(2), 217-220, 1981.

Shuiskaya, F.K., Yu. I. Galperin, A.A., Serov, N.V. Baranets, Yu.V. Kushnerevsky, G.V. Vasil'ev, S.A. Pulinets, M.D. Fligel and V.V. Selegey, Resonant heating of the ionospheric plasma by powerful radiopulses aboard the Intercosmos-19 and Cosmos-1809 satellites, *Planet. Space Sci.*, *38*(2), 173-180, 1990.

Trotignon, J.G., P. M. E. Décréau, J. L. Rauch, E. Le Guirriec, P. Canu, and F. Darrouzet. The Whisper Relaxation Sounder Onboard Cluster: A Powerful Tool for Space Plasma Diagnosis around the Earth, *Cosmic Res. 41*(4), 345-3348, 2004.

Wakai, N., "Operation and experimental results of the Ionosphere sounding satellite ISS-b", *Acta Astronautica, 7*, 999-1020, 1980.

Yau, A.W., H.G. James and W. Liu, The Canadian Enhanced Polar Outflow Probe. (e-POP) Mission in ILWS, *Advances in Space Research*, *38*(8), 1870-1877, DOI:10.1016/j.asr.2005.01.058, 2006.

2.3   THE SuperDARN RADAR NETWORK

(William A. Bristow, Geophysical Institute University of Alaska Fairbanks)

The Super Dual Auroral Radar Network, SuperDARN, [Greenwald *et al*., 1995] is the product of more than twenty years of research and development which began in 1983 with the construction of an HF radar at Goose Bay, Labrador. In the intervening years, the HF radar technique has had a major impact on high-latitude ionospheric research in both the aeronomy and magnetospheric physics communities. During that time, SuperDARN has evolved from a single independent radar to a network of ten operational radars in the northern hemisphere and seven operational radars in the southern hemisphere. A plan view of the current coverage of the northern network is shown in Figure 2-5. With the addition of the Alaskan radars in 2000 and 2001, there is continuous SuperDARN coverage extending form northern Scandinavia westward to Siberia with extensive conjugate coverage in the southern hemisphere. In the standard mode of operation, each radar measures the line-of-sight plasma velocity at 75 ranges along each of 16 beam directions covering an area of about 3500 km in range and about 56 degrees in azimuth. The radars operate continuously, 24 hours per day and 7 days per week and deliver images of the high-latitude convection pattern with a one to two-minute time resolution.
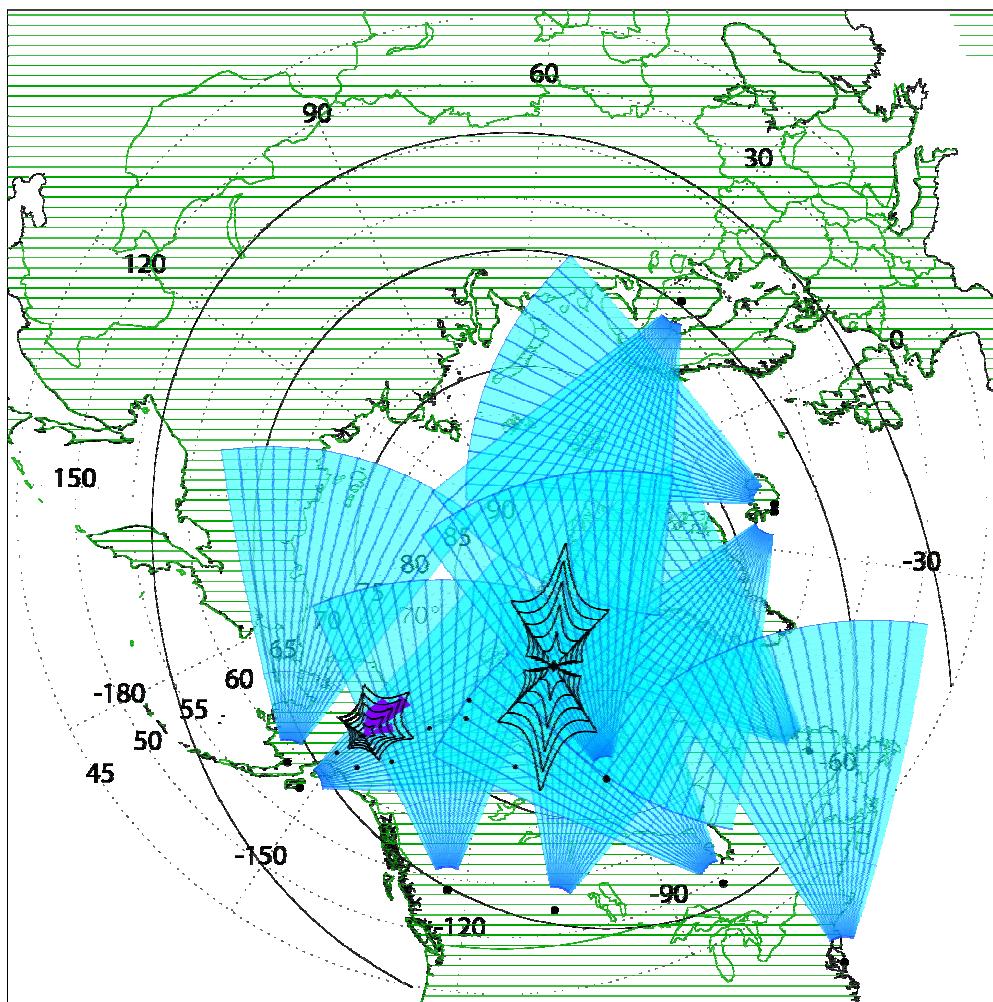
Figure 2-5: Fields of View of the Northern SuperDARN Radars.

The main purpose of SuperDARN is the instantaneous mapping of the ionospheric convection pattern. By providing direct measurements of convection velocities over large regions, SuperDARN provides the best means available to derive ionospheric convection patterns. Prior to the development of SuperDARN, our knowledge of convection was based primarily upon conceptual models, simulations, and empirical models that were developed from observations accumulated over time. Convection data that have been available include: single point radar observations of plasma flows, low altitude satellite observations of plasma drifts, high-altitude satellite observations of electric fields, and observations of magnetic perturbations by satellite-based and ground-based magnetometers. Each of the previously available techniques has characteristics that may cause the derived pattern to differ from the actual pattern at any given moment. SuperDARN provides direct, instantaneous, measurement of convection and hence, avoids many of the drawbacks of other techniques.

Each radar consists of a phased-array antenna system, sixteen transmitters, a receiver, a phasing matrix, and a computer system for control and data logging. The antenna system consists of two rows of log periodic antennas covering the frequency range from 8 MHz to 20 MHz. One row, of sixteen antennas, is used for transmission and reception, the second row, of four antennas, is used for reception only. Both arrays are steered into sixteen viewing directions by an electronic phasing matrix. The two rows of antennas are separated along their array normal direction and are used as an interferometer to derive angle-of-arrival information from the received signals.

Data from the network can be accessed in real time through the SuperDARN page. Follow real-time data link and then the Java applet link and read the instructions on how to view the data. Figure 2-6 is an example display from the Kodiak radar. This image is from about 1830 UT on January 27, 2000. It was a relatively active interval when Alaska was in the dawn sector. The gray region represents the ground-scattered portion of the signal, that is, signal that is transmitted from the radar, reflects obliquely off of the lower portion of the F-region ionosphere down to the ground where it scatters back along the same path to the radar. The ground scatter is useful for obtaining the density structure of the bottom-side F-region. The colored regions represent scatter from decameter-scale plasma irregularities. These irregularities are in the F-region and move with the bulk plasma velocity. The color scale corresponds to the component of the plasma velocity along the radar line of site. The red and yellow colors correspond to velocity away from the radar, while blue colors correspond to velocity toward the radar.
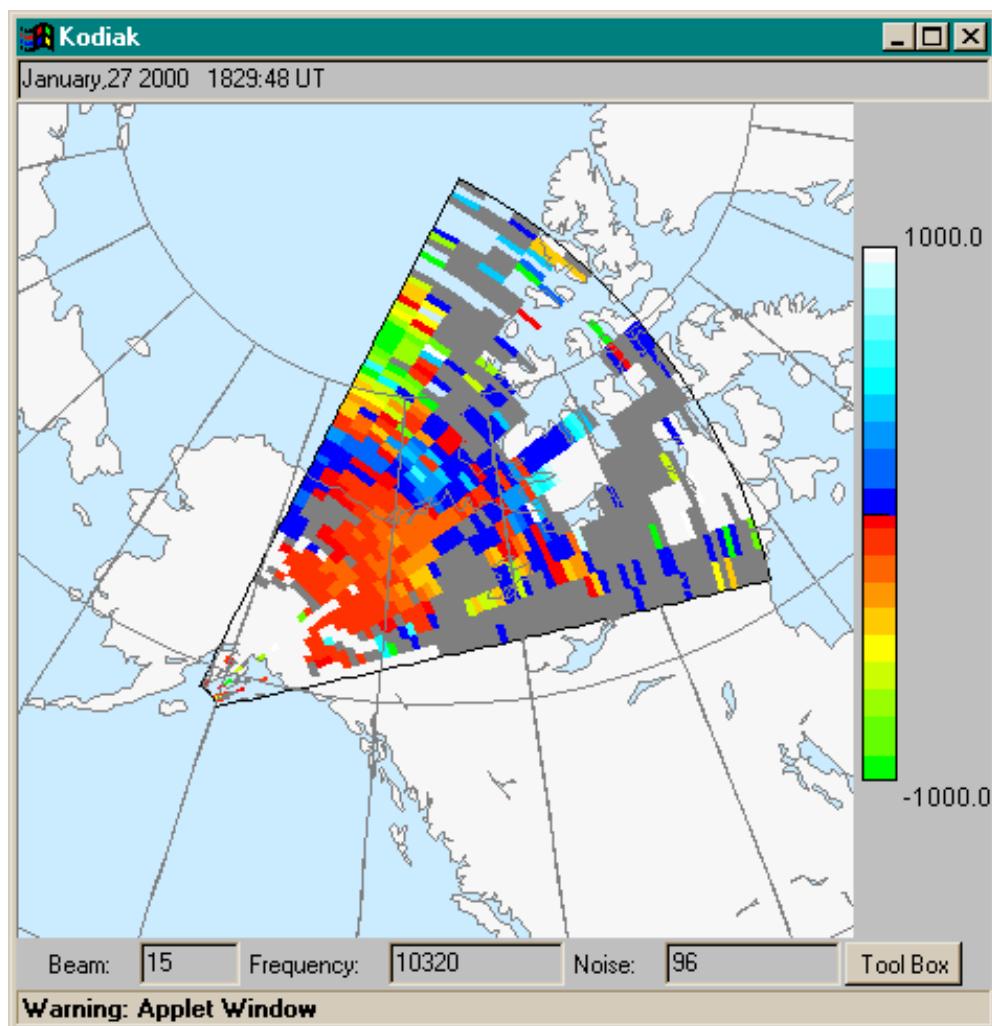
Figure 2-6: Example of Velocity Observations from the Kodiak SuperDARN Radar.

In addition, the data are being assimilated into the northern-hemisphere convection mapping, which also can be viewed through the real-time data link. Below is and example convection map from the same interval as the above image of the Kodiak field of view.
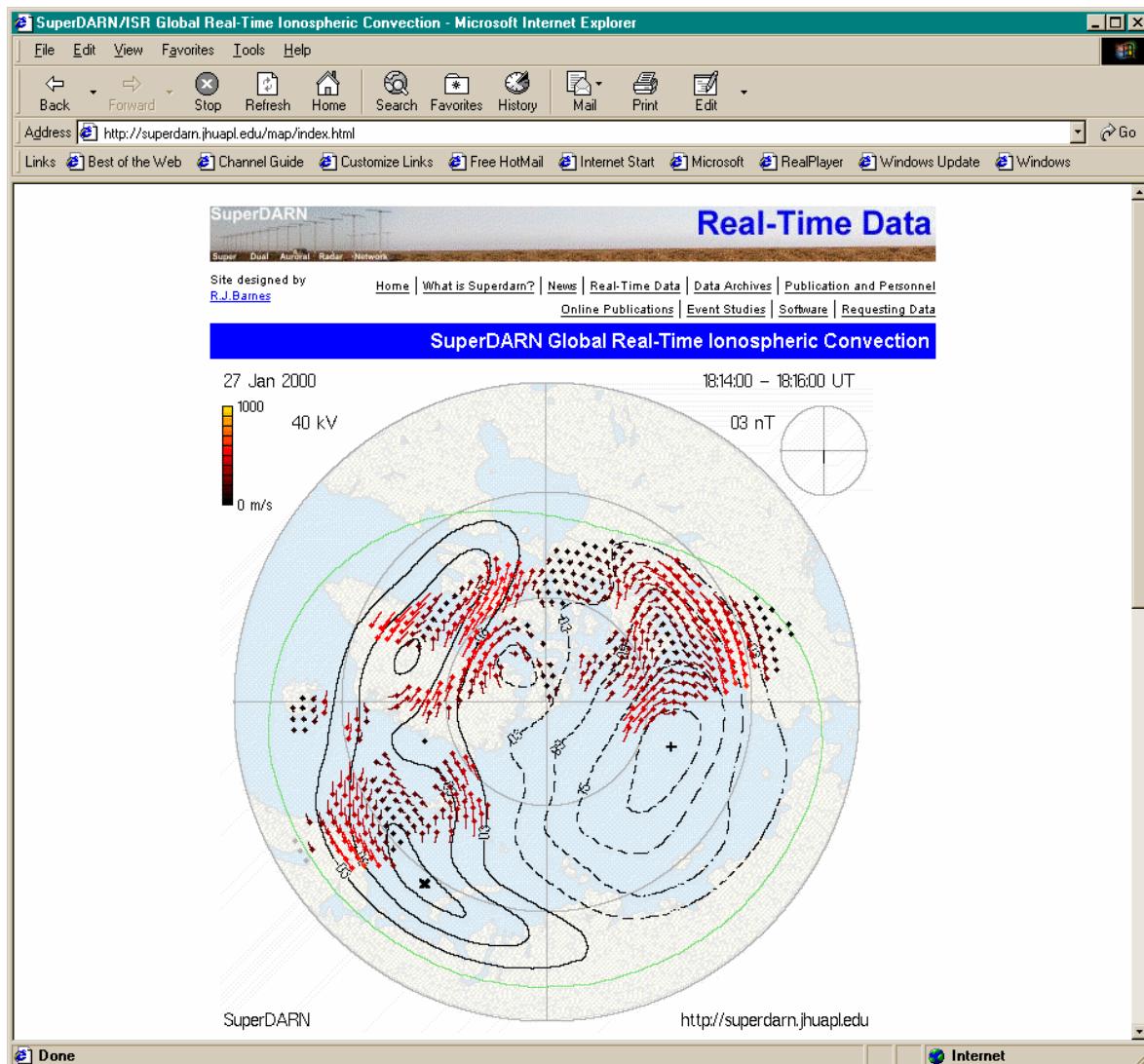
Figure 2-7: Example Convection Pattern from the Real-Time Display on the JHUAPL Web Page.

In addition to convection pattern measurements, SuperDARN has been used for many other types of observations. A listing of publications that rely on data from SuperDARN can be found on the SuperDARN web page at the Johns Hopkins University Applied Physics Laboratory: http://superdarn.jhuapl.edu.

Radio propagation over Alaska is one example of studies enabled by SuperDARN. The work carried out in this area has resulted in development of a measurement technique for estimating HF communication parameters [Hughes et al, 2002], which has had the side benefit of improving operation of the network. The principle used for estimating HF communications parameters with an HF radar is simple. Under common conditions, a portion of the signal transmitted from the radar reflects from the ionosphere towards the ground. Upon striking the ground, some of the signal is scattered back in the direction from which it was incident and returns to the radar. This signal is termed ground scatter, and is easily distinguished from the scatter from ionospheric plasma irregularities. Reflection of a ray from the ionosphere depends on the ionospheric plasma density, the frequency of the signal, and the angle of incidence of the ray. By determining the critical angle below which reflection occurs as a function of frequency, it is possible to map the F-region critical frequency for vertical incidence (f_oF_2), which directly translates to the peak plasma density. The experiment technique developed for this observation used spare

time available in the normal SuperDARN measurement cycle to rapidly step through frequencies and observe the ground scatter. This technique provides estimates of the F-region peak density over a range of distances along each beam direction of the radar. Figure 2-8 shows a map of these f_oF_2 observations. The technique has the side benefit that while observing ground scatter, it is also possible to determine the optimum frequency for observing ionospheric scatter. A second algorithm exploits this to automatically select the radar operating frequency, which has resulted in a great increase in the amount of ionospheric scatter observed by the radars.
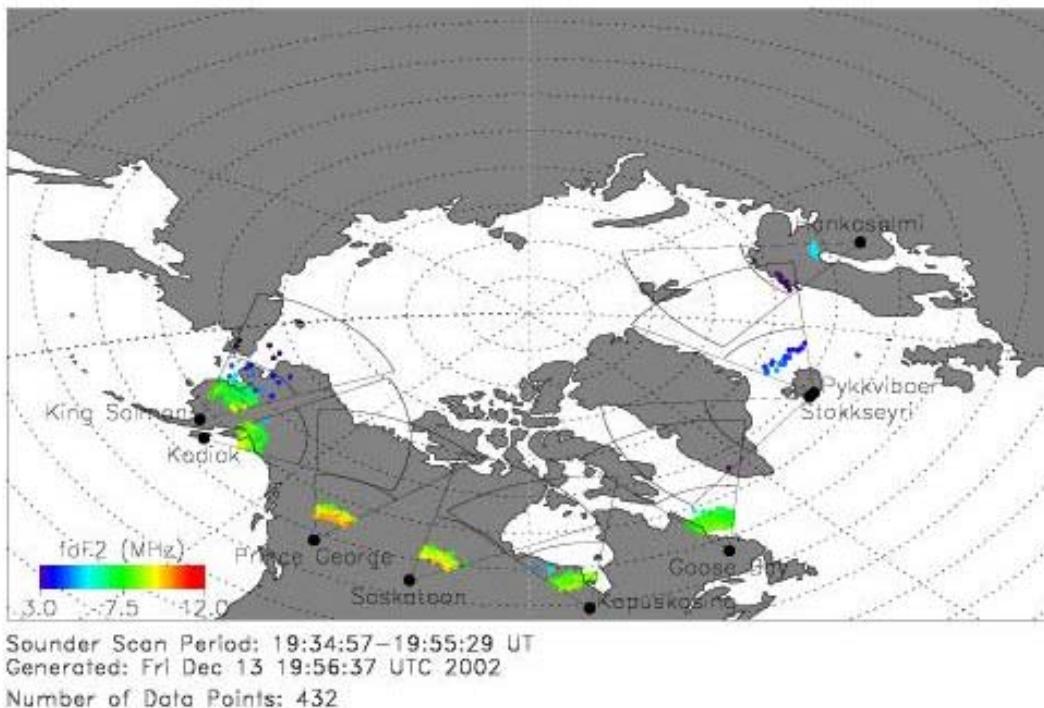
Figure 2-8: f_oF_2 Estimates from the SuperDARN Network.

SuperDARN Contributions to the HAARP/HIPAS programs have been significant. Primarily, work has focused on HAARP, which is a DOD operated facility for radio frequency heating of the ionospheric plasma. In heating the plasma, various instabilities generate plasma irregularities similar in character to the naturally occurring irregularities, which are the primary target of the Kodiak radar. Because the artificial irregularities reliably generate strong scatter, the radar has become one of the primary diagnostics for HAARP heating experiments. The radar at Kodiak and the group at UAF have supported every HAARP experiment campaign since the radar was completed. The primary focus of the UAF group has been determining the characteristics of the irregularities including their formation time, decay time, threshold for generation, effect of the angle between the RF heating signal and the Earth's magnetic field, and the frequency of the heating signal. Recently, installation of a digital receiver enabled experiments to search for reflections from various other plasma wave modes. The search was indeed successful. A variety of lines were observed in the reflection spectrum with frequencies that were related to the gyro-harmonics and other characteristic plasma frequencies. One segment of the signals had distinct characteristics that could be identified as scatter from upper-hybrid waves. These never before reported observations resulted in two manuscripts [Hughes et al, 2003; Hughes et al., 2004], and will be the seeds for much further research.

In the studies of magnetospheric field-aligned currents, global-scale convection data from the SuperDARN network were combined with ionospheric conductance patterns estimated from ultraviolet images of

auroral emissions observed by the Polar Ultra Violet Imager (UVI). Conductance estimates were determined by observing the intensities and the ratio of the upper and lower Lymon-Birge-Hopfield bands. Polar UVI provided these images with about a 1-minute cadence during some periods, which allowed studies of the simultaneously evolving conductance and convection patterns. Combining the conductance and the convection provided estimates of the currents flowing perpendicular to the Earth's magnetic field from which it is possible to infer the field-aligned currents by realizing that in regions where the perpendicular currents diverge, currents must flow along the magnetic field. These field-aligned currents are the primary mechanism by which energy flows from the magnetosphere into the ionosphere and upper atmosphere. The Bristow et al. [2001] study examined the evolution of currents in the high-latitude and polar regions as solar-wind conditions changed and during one substorm period. The strongest currents observed in the day side were transients in the cusp region, while on the night side, strong currents were observed to follow the head of the substorm expansion bulge.

Observing substorm convection using SuperDARN was a direct follow on to the field-aligned current studies. Again, observations from Polar UVI were combined with convection observations, with the addition of ground-based magnetometer observations, and observations from all-sky imagers and meridian scanning photometers. A number of substorm intervals were examined to determine general characteristics in the relationship between the plasma flow, the optical emissions, and the currents. In the initial study [Bristow et al., 2001] it was found that the global-scale convection pattern evolved steadily in the substorm growth phase to enhance the shear in the pre-midnight convection resulting in evolution of what magnetometers observe as the Harang discontinuity. The discontinuity moved to lower latitudes as the shear increased, and extended across the midnight meridian. After a period of time, zonal convection near midnight was observed to increase rapidly to high velocities. Then, as the auroral emissions increased, the convection velocity decreased, the shear in the pattern decreased, and flow near midnight rotated from zonal to nearly meridional out of the polar cap. The second study [Bristow et al., 2003] focused on the detailed evolution and sequence of events during the late growth phase and expansion phase of substorms.

A definite sequence of events was identified in which the first evidence of expansion was a surge of convection, followed by an increase in auroral luminosity, and finally followed by the negative magnetometer bay that is often used to identify substorms. In addition, it demonstrated that there was no evidence for magnetic merging until the time of the negative magnetometer bay. The magnetometer bay also coincided with relaxation of the convection shear, and rotation of flow to being directed out of the polar cap. Figure 2-9 illustrates the merging of the radar data with ground-based all-sky imager observations of the aurora. The correspondence between the observed sheer in the flow with the location of auroral arcs is phenomenal.
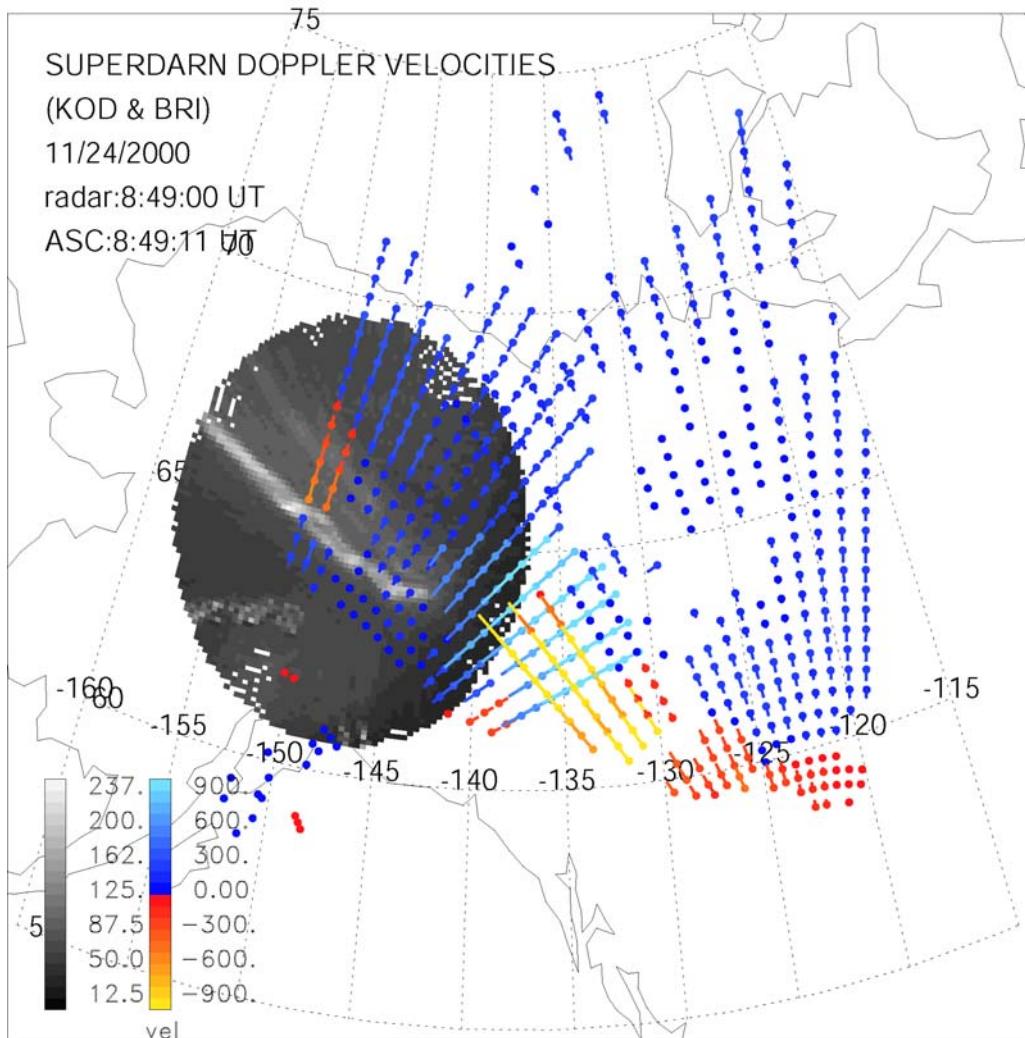
Figure 2-9: Superposed Plasma Velocity Observations and Auroral Luminosity Observations.

All of the topics discussed so far relate to properties of the ionospheric plasma, or use reflections from the ionospheric plasma to observe magnetospheric phenomena. In addition to these topics, the SuperDARN group has exploited scatter from ablating meteors to observe properties of the neutral atmosphere winds in the mesosphere. During the interval reported here, two masters' theses [Wei, 2000; Parris, 2003] have reported various aspects of this topic. Wei [2000] developed an algorithm to separate meteor scatter from other types of scatter in the SuperDARN data set. He was able to examine mesospheric winds at an assumed altitude of about 95 km, and was able to extract the various tidal modes and planetary waves at polar latitudes, and carry out a preliminary analysis of their seasonal behavior. Parris [2003] carried these investigations further and developed an experimental technique for extracting the meteor scatter as a function of altitude. This required replacing the analog receiver at Kodiak with a digital receiver that allowed over sampling of the received signal and multiple data channels. Having multiple channels resulted in two different data streams, which allowed the over-sampled data to be independent of the normal SuperDARN data. The nature of the meteor signal allows treatment as a hard target, providing greater spatial resolution from long duration signals. Exploiting this has provided altitude profiles in the range from about 70 km altitude to about 110 km altitude.

2.3.1 References

Bristow, W. A., D. Lummerzheim, Determination of field-aligned currents using the Super Dual Auroral Radar Network and the UVI ultraviolet imager, J. Geophys. Res. 106, 18,577-18,588, 2001.

Bristow, W. A., A. Otto, and D. Lummerzheim, Substorm convection patterns observed by the Super Dual Auroral Radar network, J. Geophys. Res. 106, 24,593-24,609, 2001.

Bristow, W. A., G. J. Sofko, H. C. Stenbaek-Nielsen, S. Wei, D. Lummerzheim, A. Otto, Detailed analysis of substorm observations using SuperDARN, UVI, ground-based magnetometers, and all-sky imagers, J. Geophys. Res. 108, A3, 1124, doi:10.1029/2002JA009242, 2003.

Greenwald, R. A., et al., DARN/SuperDARN: A global view of high-latitude convection, Space Sci. Rev., 71, 763-796, 1995.

Hughes, J. M., W. A. Bristow, R. A. Greenwald, R. J. Barnes, Determining characteristics of HF communications links using SuperDARN, Ann. Geophys., 20, 1023-1030, 2002.

Hughes, J. M., W. A. Bristow, R. T. Parris, E. Lundell, SuperDARN observations of ionospheric heater-induced upper hybrid waves, Geophys. Res. Lett., 30, doi:10.129/2003GL018772, 2003.

Hughes, J. M.; Bristow, W. A.; Parris, R. T. SuperDARN observations of spectral enhancements excited during an ionospheric heating experiment, Geophys. Res. Lett., Vol. 31, No. 8, L08808 10.1029/2004GL019613, 2004.

Parris, R. T., Design and implementation of a meteor tracking retrofit system for the HF radar at Kodiak Island, Alaska, Unpublished M.S. EE thesis, University of Alaska Fairbanks, 2003.

Wei, D., Meteor Wind Wave Analysis of SuperDARN Observations, Unpublished M.S. EE thesis, University of Alaska Fairbanks, 2000.

Chapter 3 – MAPPING

3.1 CREATING MAPS OF IONOSPHERIC ELECTRON DENSITY TO SUPPORT COMMUNICATION, SURVEILLANCE AND NAVIGATION SYSTEMS

(Cathryn Mitchell, Department of Electronic and Electrical Engineering, University of Bath, Bath, UK)

3.1.1 Introduction

Currently there are a number of different methods to produce estimates of the electron density in the ionosphere. Model-based approaches range from empirical models (e.g. IRI, Bilitza, 2001), through to methods that combine such models with real-time data (e.g. Bust et al., 2004) and finally approaches that constrain a physical model with measurements (e.g. Hajj et al., 2004). For well-instrumented regions of the Earth it is possible to design algorithms that use mathematical constraints and measurements to produce electron-density images without including ionospheric models (Spencer and Mitchell, 2001; Mitchell and Spencer, 2003). For global solutions of the electron density either model-assisted approaches or interpolation methods (e.g. solar-fixed persistence) are required to fill in large data-sparse regions of the Southern Hemisphere, in particular over the oceans.

This paper focuses on the development of the Multi-Instrument Data Analysis System (MIDAS) algorithm. Of all the approaches summarised above MIDAS is probably the most 'mathematical and data driven' and least 'model reliant'. This approach was taken in the development of MIDAS for a number of reasons. Firstly, it was apparent an early stage in the development (1999) that the number of available data sources for the ionosphere was increasing, mainly because of the increased use of GPS for scientific studies in other fields (mainly geodesy). In addition, it was also clear that much of the available data already being recorded was not being systematically utilized for ionospheric imaging (e.g. navigation data on low Earth orbit, LEO, satellites; Superdarn; ad-hoc HF links), either because of a lack of central collation or because of the required nonlinear inversion techniques had not yet been developed. Finally, the strength of minimising the model-reliance in the MIDAS approach is that in data-rich regions images that can show actual behaviour of the ionosphere at a given instant and not a 'merged' model-data solution. This also makes MIDAS a suitable technique for producing statistical results that could be used to produce new ionospheric models with improved ability match actual events.

In this paper the development of MIDAS is described. Examples of results for global and local ionospheric specification are given. Finally the future directions and required modifications to the technique for specific systems uses are discussed.

3.1.2 MIDAS Development

The MIDAS algorithms are based upon the oceanographic imaging techniques of Munk and Wunsch (1979); first applied to imaging 2D slices of the ionosphere by Fremouw et al. (1992). This was conventional tomographic imaging where the line integral data were approximated into a plane and inverted to reveal the electron density. MIDAS (version 1.0) is a linear inversion algorithm that can ingest any line-integral data such as GPS-ground or GPS-LEO differential-phase data (Yin et al. 2005) or inverted ionograms.

The development of the MIDAS algorithms was driven by the success of two-dimensional ionospheric tomography for both scientific and applications work (e.g. Mitchell et al., 1995; Mitchell et al., 1998 and Rogers et al., 2001). Initially it was thought that GPS data could only be used for simple constant altitude shell mapping and early work concentrated in this area. Comparisons between shell mapping and full 3-D imaging can be found in Meggs et al., 2004. In fact, the step from 2-D shell mapping to full 3-D imaging

is a matter of including mathematical concepts from tomography (Mitchell, 2002; Mitchell and Spencer, 2003). It was apparent that the differential phase-technique could only work with a time dependent algorithm, because the ionosphere changes while the satellite moves. Alternatively, pre-calibration of the differential code data could allow a time independent approach. Extensive simulation studies were used in the development of the techniques for the equatorial region (Materassi et al., 2005) and the mid-latitudes (Meggs et al., 2004). The imaging has recently been applied to the sparse-data region of South Africa (Cilliers et al., 2004).

The first stage of the MIDAS algorithm (Mitchell and Spencer, 2003) inversion is to set up a three-dimensional grid of voxels, each bounded in latitude, longitude and altitude, and to compute the length of each element of a satellite-to-receiver signal propagation path though each intersected voxel. The unknown electron concentration, $\mathbf{x_a}$, is defined to be constant within each voxel and contained in the column vector. The problem may now be expressed as:

$$\mathbf{y} \; = \; \tilde{\mathbf{H}} \mathbf{x_a}$$

where the matrix $\mathbf{H}$ transforms the electron density to the form and location of the observations and $\mathbf{y}$ are the observed TECs. It should be noted that the inversion uses relative differential phase observations. Thus appropriate lines of the matrix are differenced such that measurements along continuous satellite-receiver arcs are taken relative to a certain reference measurement within that arc.

The inversion cannot be performed directly so a mapping matrix, $\mathbf{X}$, is used to transform the problem. This results in a situation where the unknowns are coefficients of orthonormal basis functions, the combination of which will give the final image of electron concentration. The basis functions ($\mathbf{X}$) can be generated using a spherical harmonic expansion to represent the horizontal variation and empirical ortho-normal functions (EOFs) for the radial variation in electron concentration. The spherical harmonics provide a flexible basis to determine the horizontal distribution of ionization, which should be well defined by the measurements. The EOFs form a constraint to the vertical profile, only allowing a certain range of possible solutions. This is now expressed mathematically as:

$$\mathbf{y} = \; \tilde{\mathbf{H}} \tilde{\mathbf{X}} \tilde{\mathbf{W}}$$

where the matrix $\mathbf{X}$ contains the basis functions. $\mathbf{HX}$ now represents the set of TEC data that are formed by integration through the set of models. Applying singular value decomposition the inversion can be performed such that:

$$\tilde{\mathbf{W}} = \; (\tilde{\mathbf{H}} \tilde{\mathbf{X}})^{-1} \mathbf{y}$$

and the solution to the inverse problem is then given by:

$$\mathbf{x_a} = \; \tilde{\mathbf{W}} \tilde{\mathbf{X}}$$

The algorithm can be extended into a time-dependent inversion by incorporating a priori information about the evolution of the electron concentration during a specified period of time. Assuming that the change in electron concentration within a voxel with time is linear, then it is possible to write the same system of equations to solve for the change in the relative contributions of each basis function.

Recent developments to MIDAS have improved the ability to image in the data-sparse regions of the polar cap by incorporating a motion model into the imaging there. Specifically, a Kalman-filter approach is used to incorporate the Weimer model of the plasma convection into the imaging, thus allowing for the continuity of features as they move across data gaps (Spencer and Mitchell, 2007). It has also incorporated a new grid-rotation has been implemented that provides a solution to the problems caused to the imaging by convergence of the spherical coordinates at the pole.

3.1.3 Example Results from MIDAS

The date of the 30th October 2003 has been chosen to illustrate the imaging. All of the images are at 1800 UT. In the demonstrations here only ground-based GPS data are used, although other measurements from ionosondes and LEO satellites were available and could be incorporated into the MIDAS images if required. This major storm exhibited unusual features in the ionosphere and serves well to demonstrate the need for measurements rather than reliance on statistical models. In order to demonstrate this further the IRI model has been used to provide a comparison to the MIDAS images.

Figure 3-1 shows the locations of the receivers used for the example images. Figure 3-1a is for the Arctic region, Figure 3-1b and 3-1c for the whole Earth.

a

b

c
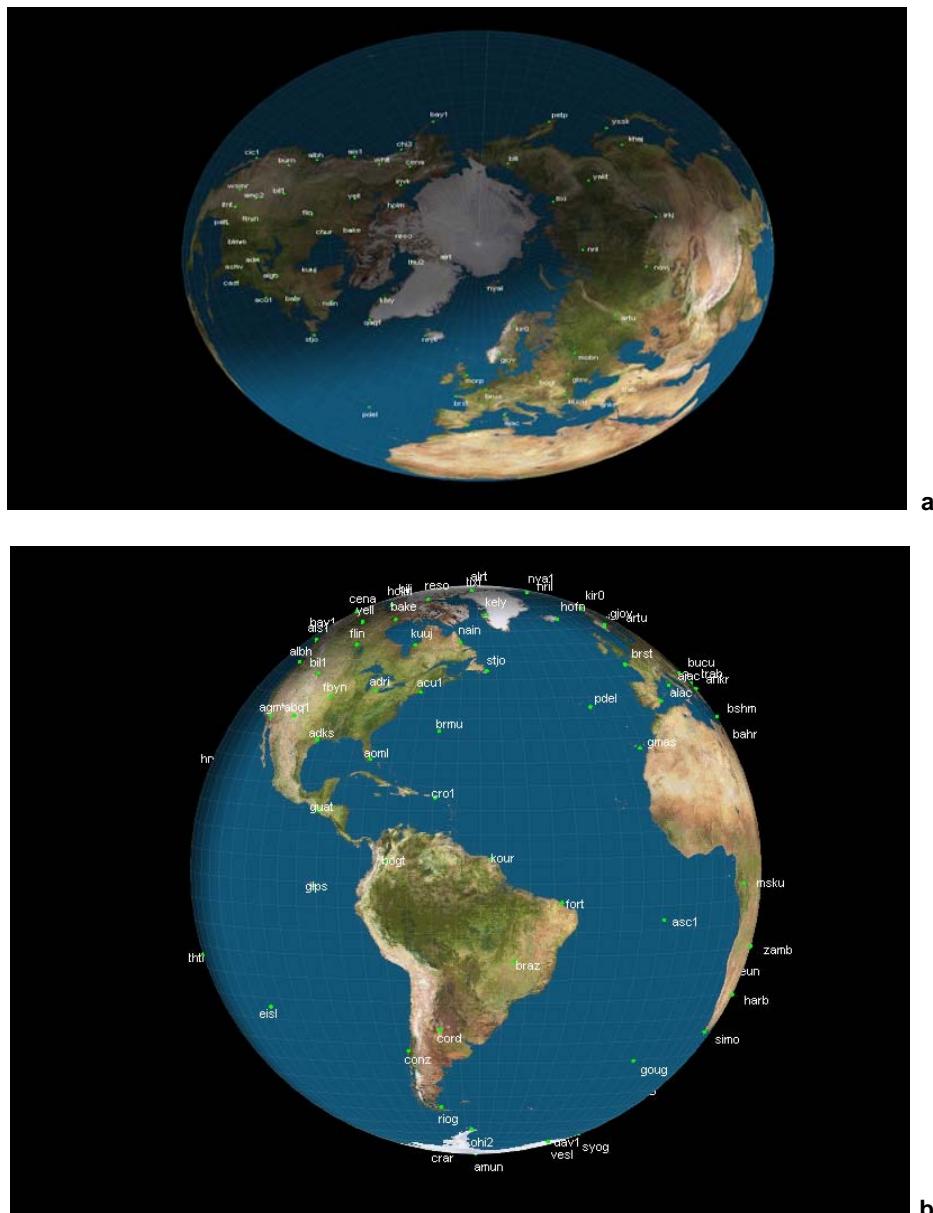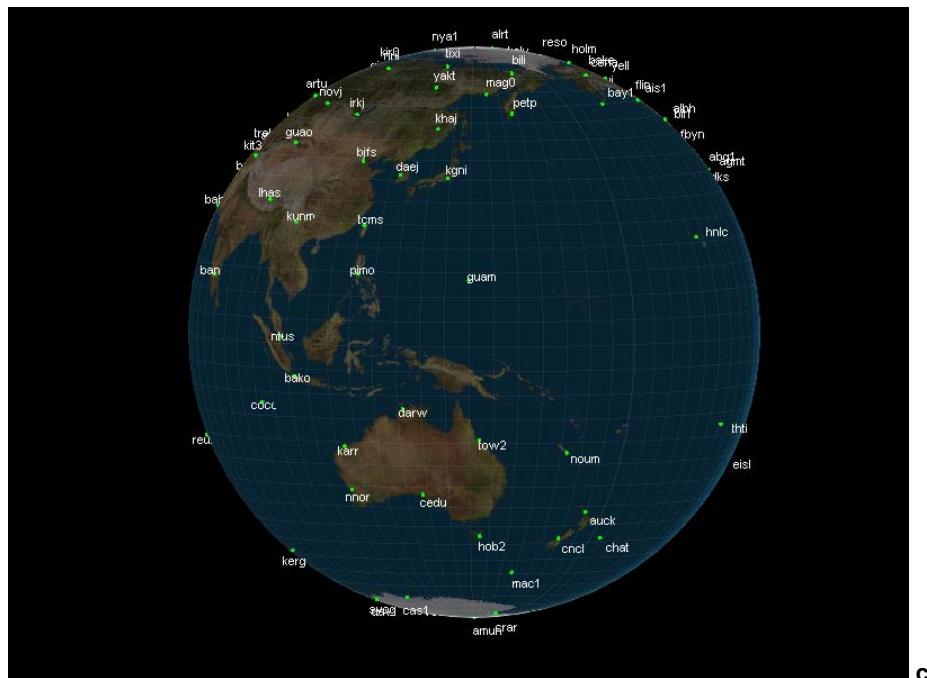
Figure 3-1: The Locations of the GPS Receivers Used for the Example Images.

Figure 3-2 shows the MIDAS images from the data coverage in Figure 3-1. The plots show isocontours of electron density (in contours from 2 to 12×10^{11} electrons per cubic metre) which allow a '3D' view of the ionosphere, revealing the altitude uplift and density enhancement regions simultaneously. Figure 3-3 shows a global view of the ionosphere from MIDAS (where the maximum TEC value is scaled to 70 TECu). Figure 3-4 shows the corresponding IRI model results, showing that, as expected, the statistical model cannot replicate such extreme events.
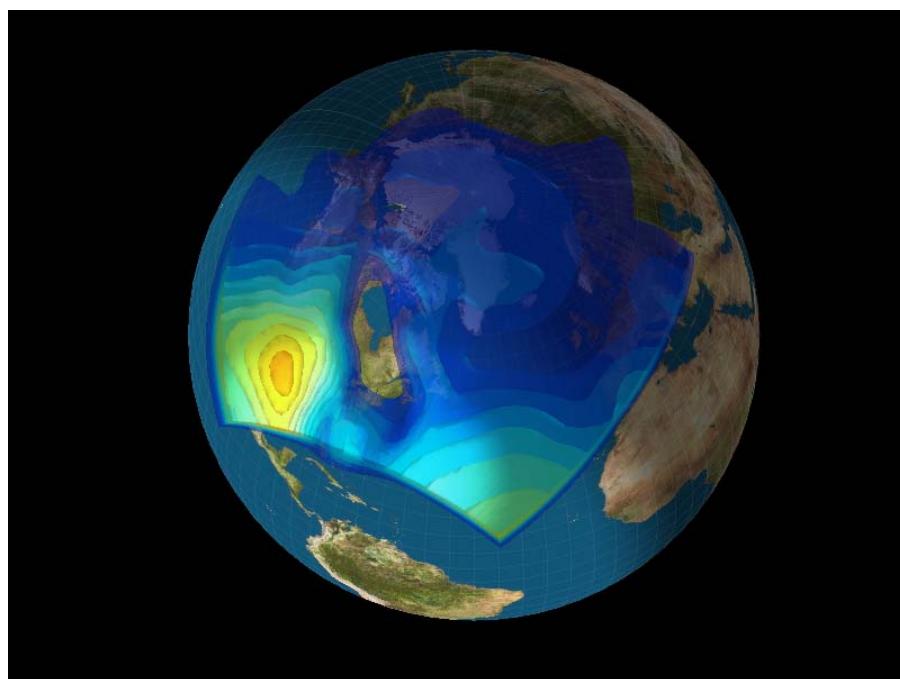
Figure 3-2: Contours of TEC Over the Polar Region from MIDAS.
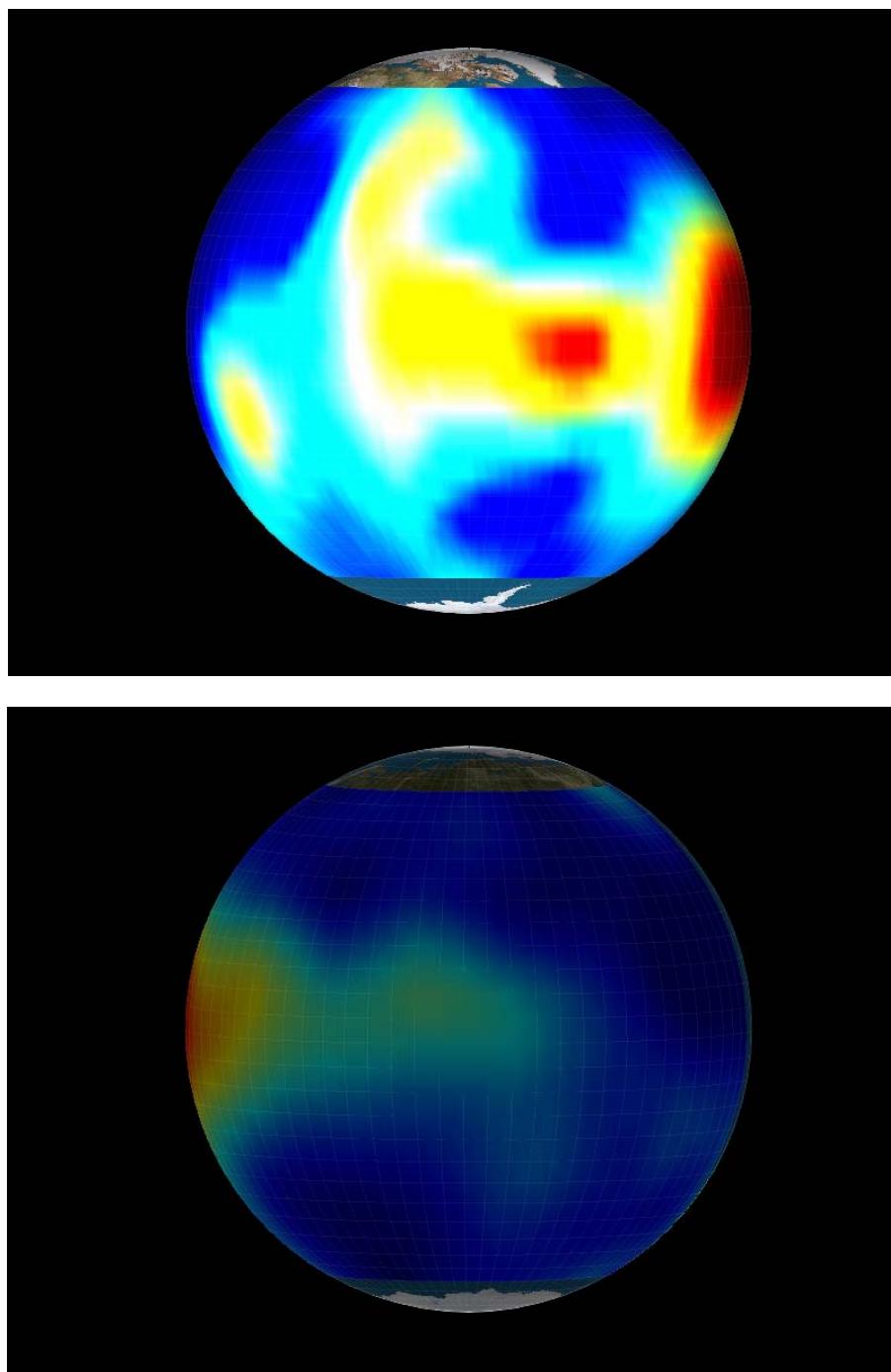
Figure 3-3: Contours of TEC Over Low-Auroral Latitudes from MIDAS.
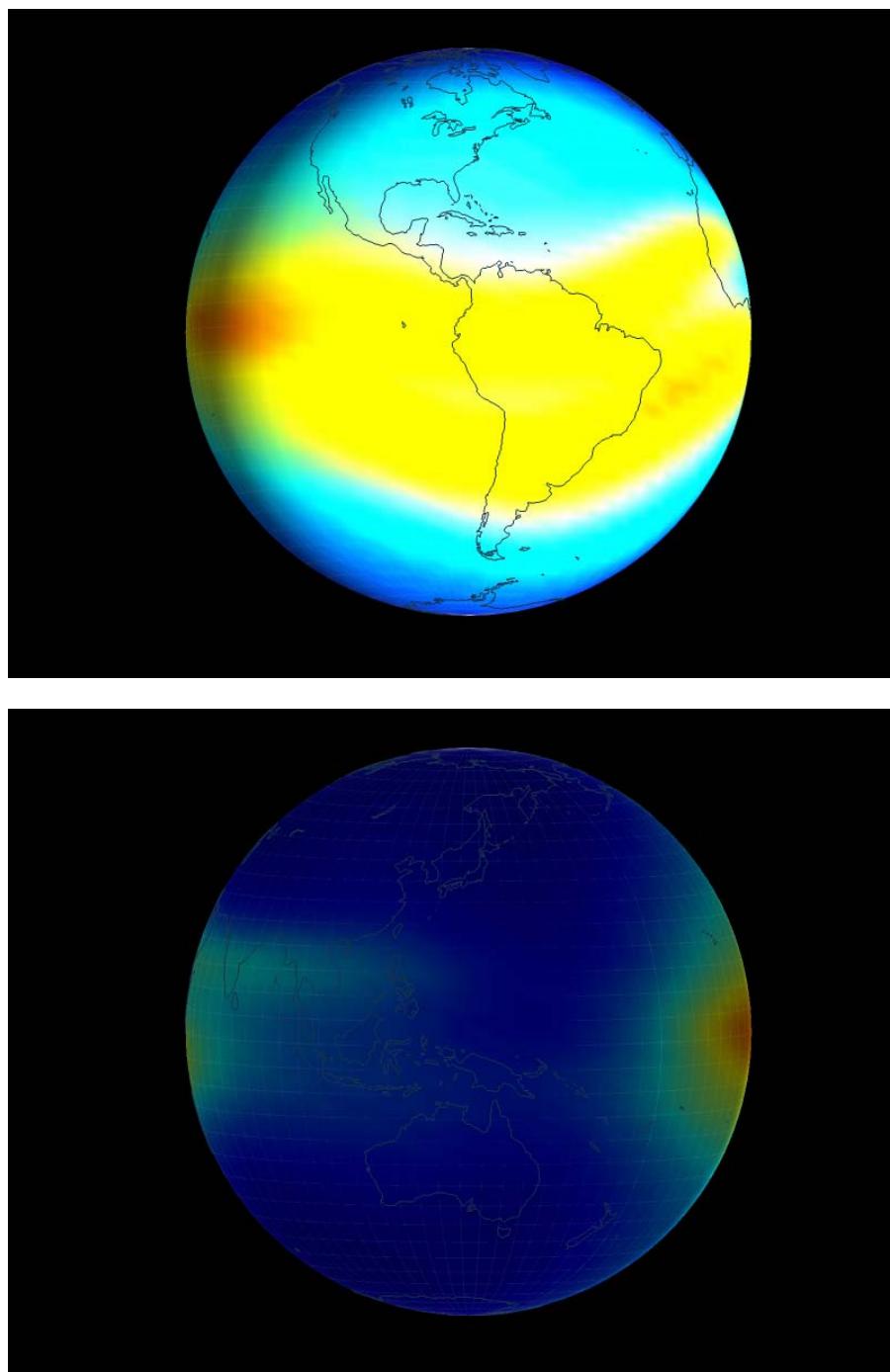
Figure 3-4: Contours of TEC Over Low-Auroral Latitudes from IRI.

3.1.4   Short-Term Forecasting and Verifications

The accuracy of the imaging technique has been examined thorough a series of simulation studies. GPS provides two different quantities that relate to the ionospheric total electron content (TEC) – the differential phase advance and the differential group delay. The phase of the signal is a highly-accurate measurement that gives a relative change in the TEC whereas the group delay is subject to stable biases that are in the largest component of their values specific to a given receiver or satellite. This dual measurement system

allows the assessment of imaging accuracy, by using only the phase data to make the images the code delay data can be reserved for image accuracy assessment. Thus through both simulation and experiment the accuracy of the imaging technique in terms of TEC mapping has been assessed (Dear et al., 2006).

The vertical profile of electron density obtained from ionospheric imaging using only ground-based data has also been assessed. A new approach was developed whereby the peak density and scale heights from ionosonde data were used to assist the reconstruction. This is particularly applicable to HF applications where the vertical profile is critical to the outcome.

Case study verifications of MIDAS have also been performed, for example in comparisons with incoherent scatter data the trough has been imaged (Meggs et al., 2005) verifications have been made showing the uplift of the ionosphere during a storm (Yin et al., 2004).

MIDAS has a short-term forecasting capability built in (up to 12 hours ahead). The forecasting algorithms are based on pattern recognition approaches. Various methods for forecasting the daytime TEC over large-regions were explored. It was found that methods that worked well for isolated locations did not translate to wide-area mapping and could produce spurious artefacts. The preferred approach involved the use of a climatological model coupled to the imaging. This was shown to be applicable even during disturbed conditions. For the night-time, the main trough was the dominant feature and its movement and density changes necessitated the use of feature recognition coupled into the forecasting (Dear et al., 2007).

MIDAS is currently being run in a real-time and forecasting mode both over the UK and over Southern Europe, and is being tested for single-frequency GPS ionospheric delay corrections.

3.1.5 Summary and Discussion

One method to produce maps of the electron density, MIDAS, has been described here. The method is the least model-reliant of a family of techniques that range from tomographic imaging to data assimilation into a physical model.

MIDAS can be used to image the ionosphere either on a low-resolution global scale (2 degree voxels) or locally at higher resolutions (typically 0.25 degree voxels). The vertical size of the voxels ranges from 5 km to 50 km. The decisions about which mode to run in depend on many factors that are outlined below.

For applications where the bulk electron density and TEC (large-scale greater than 100km structures) are the main interest the technique performs equally well in high or low-resolution mode and there is not need to operate at high resolution. Nearly all of the testing of MIDAS that has been done is essentially for this type of imaging. Examples for this application include Faraday rotation estimates for SAR and single-frequency GPS delay mapping.

For higher resolution imaging, for example to examine localised gradients in the ionosphere, there is little chance to find independent instruments to compare with. Case study comparisons with ISRs show that the gradient on the southern wall of the main trough are well replicated (Meggs et al., 2005). Larger sample (1 year) statistical studies of the trough have been done with MIDAS (Mitchell 2004, USAF EOARD report) and verified using Langmuir probe measurements where available. These results are of interest for in-beam synthetic aperture radar studies and relate to features on length scales some km to tens of km.

For applications requiring accurate representations of vertical gradients in the ionosphere (e.g. HF communications coverage maps) data or models containing good information about the bottomside gradients must be used into the MIDAS imaging. Examples of such data sources include oblique sounders, ionosondes and radio-occultations.

Future developments to MIDAS include a statistical tomography mode for imaging regions causing ionospheric scintillations likely to be responsible for communication and navigation outages in the equatorial region.

3.1.6   Acknowledgements

3.1.7   References

[1]   Bilitza, D., International Reference Ionosphere 2000, Radio Science, 36 (2): 261-275 Mar-Apr 2001.

[2]   Bust, G. S., T. W. Garner, T. L. Gaussiran II, Ionospheric data assimilation three-dimensional (IDA3D): A global, multisensor, electron density specification algorithm, Journal of geophysical research-space physics 109 (a11): art. No. A11312 Nov 27 2004.

[3]   Cilliers, P. J., Opperman, B. D. L., Mitchell, C. N., Spencer, P. S. J., Electron density profiles determined from tomographic reconstruction of total electron content obtained from GPS dual frequency data: First results from the South African network of dual frequency GPS receiver stations, IRI: Quantifying Ionospheric Variability Advances In Space Research 34 (9)2049-2055 2004.

[4]   Dear, R. M., Mitchell C.N., GPS interfrequency biases and total electron content errors in ionospheric imaging over Europe, RADIO SCIENCE 41 (6): Art. No. RS6007 Dec 19 2006.

[5]   Dear, R. M., Wide-area mapping and forecasting of TEC, PhD Thesis, University of Bath, 2007.

[6]   Fremouw, E.J., J.A. Secan and B.M. Howe, 1992, Application of stochastic inverse theory to ionospheric tomography, *Radio Science*, **17**, 721.

[7]   Hajj, G.A., Wilson, B.D., Wang, C., Pi X., Rosen, I.G., Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter, Radio Science 39 (1), art. no. rs1s05 Feb 24 2004.

[8]   Materassi, M. and Mitchell, C.N., Imaging of the equatorial ionosphere, Annals of Geophysics 48 (3): 477-482 Jun 2005.

[9]   Meggs, R. W., Mitchell, C.N., Spencer, P. S. J., A comparison of techniques for mapping total electron content over Europe using GPS signals Radio Science 39 (1): Art. No. RS1S10 Feb 25 2004.

[10]   Meggs, R.W., Mitchell, C.N., Howells, V.SC., Simultaneous observations of the main trough using GPS imaging and the EISCAT radar Annales Geophysicae 23 (3): 753-757 2005.

[11]   Mitchell, C.N., Spencer, P.S.J. A three-dimensional time-dependent algorithm for ionospheric imaging using GPS ANNALS OF GEOPHYSICS 46 (4): 687-696 Aug 2003.

[12] Mitchell, C.N., D.G. Jones, L. Kersley, S.E. Pryse and I.K. Walker, Imaging of field-aligned structures in the auroral ionosphere, Ann. Geophysicae, **13**, 1311-1319, 1995.

[13] Mitchell, C.N., Walker, I.K., Pryse, S.E., Kersley, L., McCrea, I.W., Jones, T.B., First complementary observations by ionospheric tomography, the EISCAT Svalbard radar and the CUTLASS HF radar, Annales Geophysicae-Atmospheres Hydrospheres And Space Sciences 16 (11): 1519-1522 Nov 1998.

[14] Mitchell, C.N. Imaging of near-Earth space plasma Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical And Engineering Sciences 360 (1801): 2805-2818 Dec 15 2002.

[15] Munk, W., and C. Wunsch, Ocean acoustic tomography: A scheme for large scale monitoring, Deep Sea Res., **26A**, 123, 1979.

[16] Rogers, N.C., Mitchell, C.N., Heaton, J.A.T., Cannon, P.S., Kersley, L., Application of radio tomographic imaging to HF oblique incidence ray tracing, RADIO SCI 36 (6): 1591-1598 Nov-Dec 2001.

[17] Spencer, P. S. J. and C. N. Mitchell, 2001, Multi-instrument Data Analysis System, Proc. Beacon Satellite Symposium, Boston.

[18] Yin, P., Mitchell, C.N., Spencer, P.S.J., Foster, J.C. Ionospheric electron concentration imaging using GPS over the USA during the storm of July 2000, Geophysical Research Letters 31 (12): Art. No. L12806 Jun 22 2004.

[19] Yin, P., Mitchell, C.N. Use of radio-occultation data for ionospheric imaging during the April 2002 disturbances Source: GPS SOLUTIONS 9 (2): 156-163 Jul 2005.

3.2 RADIO TOMOGRAPHIC INVESTIGATIONS OF PLASMA STRUCTURE IN THE HIGH-LATITUDE IONOSPHERE: CURRENT KNOWLEDGE AND OPEN QUESTIONS

(S. Eleri Pryse and Helen R. Middleton University of Wales Aberystwyth, UK)

3.2.1 Introduction

The high-latitude ionosphere is a highly structured medium, comprising electron density irregularities over a large range of scale sizes. Of particular relevance to this report are enhancements and depletions on horizontal spatial scales of tens to hundreds of kilometres. Such structures include the tongue of ionisation and patches in the polar cap, boundary blobs in the evening auroral region and the main ionisation trough. The steep density gradients associated with the features and accompanying small-scale irregularities are of particular concern to the performance of practical navigation and communications radio systems. The basic physical processes underlying the formation of the structures are reasonably well established although the relative contributions resulting from photo- and impact-ionisation, chemical loss and plasma transport in any given situation and at any particular time remain open to debate. There is a need to establish this balance before there can be a proper understanding of the complex morphology of the spatial distribution of electron density at high latitudes. The variability in the occurrence, structure and location of the observed features means that it is not yet possible to represent them accurately in physical or parameterised models of the ionosphere at high latitudes. In addition, knowledge of the effects of space-weather disturbances is inadequate and their mitigation presents particular problems to current and planned applications of navigation satellite technology.

Radio tomographic imaging has proved to be a technique that is particularly adept for the study of large-scale spatial structures in ionospheric electron density. Recent results are summarised here, derived in the main from experimental studies using the method. Some studies relate to investigations aimed at validation and development of models of the ionosphere at high latitudes that are required for the mitigation of propagation effects on applied radio systems.

3.2.2 High-Latitude Convection and Plasma Distribution

Interaction between the interplanetary magnetic field (IMF) carried by the solar wind and the geomagnetic field is the key to the convection that drives the plasma in the polar cap. Magnetic reconnection occurs when the IMF impinges on the Earth's magnetic field, with the location of reconnection being dependent on the orientation of the IMF (Dungey, 1961; Lockwood, 1998). This in turn controls the convection flow pattern of the plasma in the high-latitude ionosphere. When the IMF is directed southwards, that is with a negative B_z component, the reconnection occurs near the equatorial plane, with the resultant evolution of the flux tubes driving the familiar two-cell ionospheric convection pattern at high latitudes. There is an anti-sunward plasma drift across the polar cap, with the return flow at lower latitudes on the dawn and dusk sides. The flow patterns under other orientations of IMF become increasingly more complicated (Reiff and Burch, 1985; Cowley, 1998). A positive B_z component drives polar lobe cells, which have a flow initially sunward in the cusp region and are flanked by viscous driven convection cells on the dusk and dawn sides at lower latitudes. A B_y component results in asymmetry of the pattern around the noon-midnight meridian. In practice, the constantly changing magnitude and direction of the IMF gives rise to substantial variations in the flow pattern and consequently in the structure of the plasma that is transported through the polar cap. If the ionisation is at high altitude the chemical loss rate is slow so that long-lived structures can be transported from the dayside through the polar cap to be deposited into the auroral zone on the nightside.

Several experimental techniques have been used to measure the electron density and electron content in the polar region. The methods and results are reviewed by Crowley (1996). Complete understanding of the mechanisms responsible for the increased density in the region is still open, with both soft-particle precipitation (Weber et al., 1984) and photoionisation (Buchau et al., 1985) from sub-polar latitudes being cited. Experimental evidence suggests that dayside photoionisation from lower latitudes can be entrained into the polar convection pattern in what has become known as a tongue of ionisation (TOI) (Valladares et al.,1994). Pryse et al. (2004) proposed that the high densities originate from the equatorward side of the afternoon cell of the high-latitude convection, and supportive evidence for a latitudinally-extended feature near magnetic noon was presented by Sims et al. (2005). Several methods have been proposed for breaking up the TOI into the discrete patches that are found in the cross-polar anti-sunward flow when the B_z component of the IMF is negative (McEwen and Harris, 1996). Possible mechanisms include temporal variations in the high-latitude convection pattern in response to changes in B_y (Sojka et al., 1993), expansion and contraction in the size of the polar cap region (Anderson et al., 1988), and an increase in the rate of ionisation recombination in flow channel events (Rodger et al., 1994). Recent observations have suggested that the ionisation may be formed into patches before it enters the polar cap (Moen et al., 2006). The balance of the contributions from the different mechanisms remains an open question. An early modelling study by Robinson et al. (1985) suggested that the structures with enhanced density found in the polar cap may exit from the polar flow into the nightside sector where they become extended in longitude but restricted in latitude to form features known as boundary blobs. However, another study by Jones et al. (1997) provided convincing evidence that such localised structures in the auroral ionosphere can be produced *in situ* by soft particle precipitation.

Recent collaborative studies led by the University of Wales Aberystwyth (UWA) have investigated aspects of the effects of convection on the distribution of plasma at the high latitudes. These have been mainly case studies, but also include statistical investigations.

3.2.3 Experimental Systems

The radio tomography experiment of the University of Wales Aberystwyth is central in the investigations. It comprises a chain of ground-based satellite receivers at Ny-Ålesund (78.91ºN, 11.94ºE) and Longyearbyen (78.22ºN, 15.83ºE) on Svalbard, Bjørnøya (74.52ºN, 19.05ºE) and Tromsø (69.58ºN, 19.22ºE) on mainland Norway (Figure 3-5). The receivers monitor the phase coherent signals at 150 MHz and 400 MHz, transmitted by the satellites of the NIMS (Navy Ionospheric Monitoring System) constellation in polar orbits at altitudes of about 1100 km. The receiving systems measure the total electron content (TEC) along a large number of intersecting satellite-to-receiver ray-paths. Tomographic inversion of the TEC yields the electron density over a latitude-versus-altitude plane through the ionosphere (Pryse, 2003).
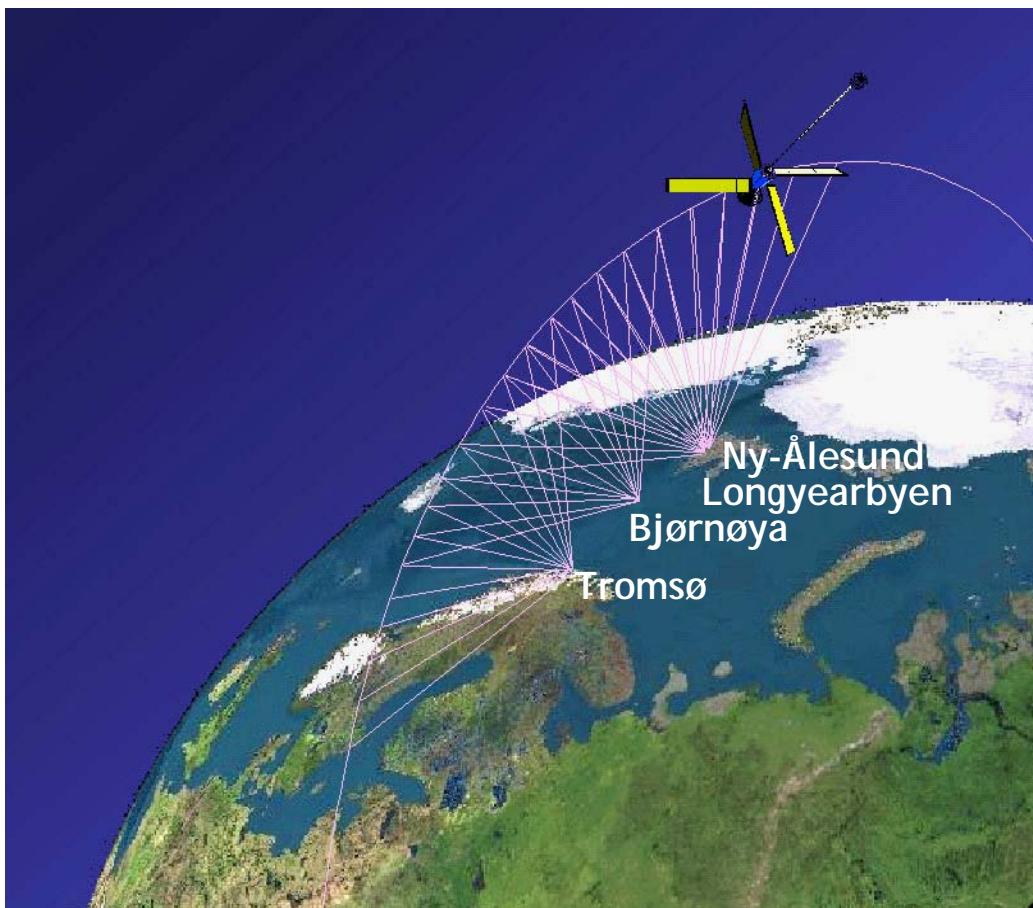
Figure 3-5: Schematic Showing the Geometry of the Radio Tomography Experiment of the University of Wales Aberystwyth in Northern Scandinavia.

The radio tomography chain forms part of the network of the International Ionospheric Tomography Community (IITC) with other high-latitude receiver chains in Alaska, Greenland, Scandinavia and Russia being operated by collaborating researchers in several countries (Figure 3-6).
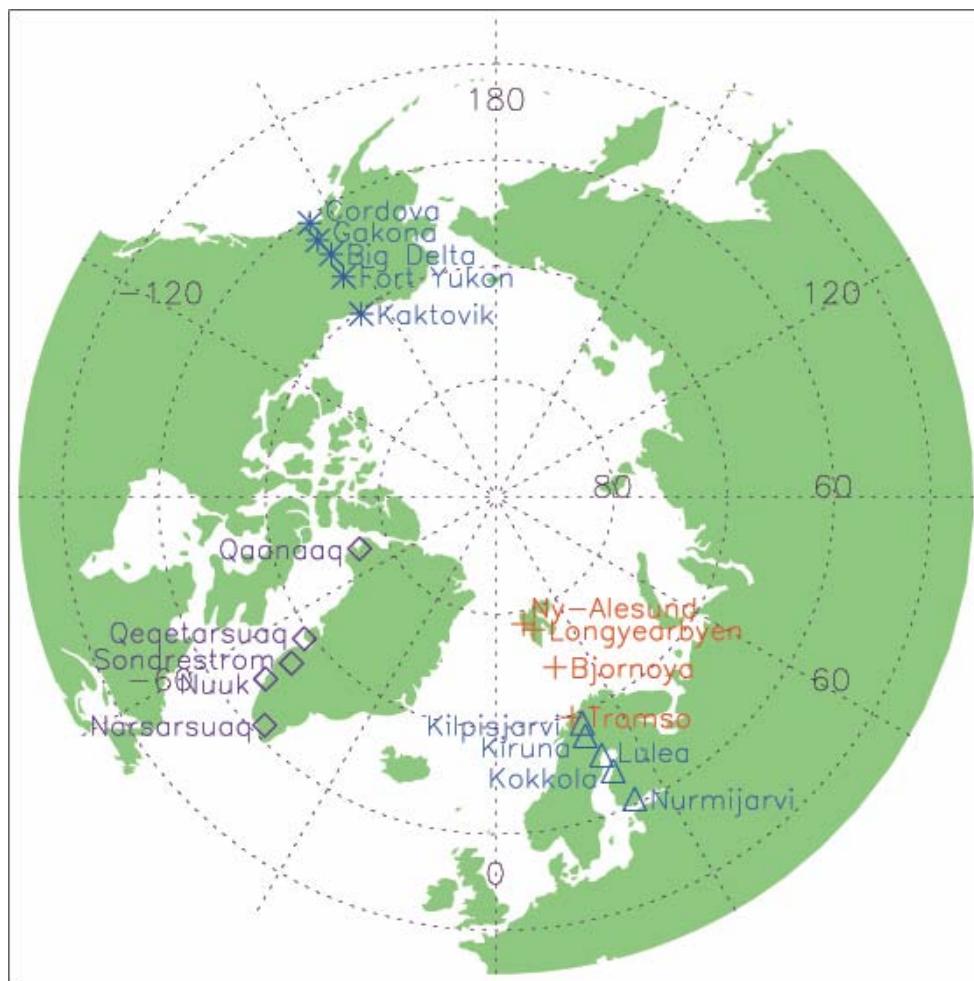
**Figure 3-6: Receiver Locations of High-Latitude Radio Tomography Chains
of the International Ionospheric Tomography Community.**

Other experimental observations used in the studies come from the European Incoherent Scatter radar facility (EISCAT) (Rishbeth and Williams, 1985) and the Super Dual Auroral Radar Network (SuperDARN) (Greenwald et al., 1995). These measurements are instrumental in efforts to identify the plasma source, to resolve spatial and temporal changes in the ionisation distribution, and to estimate the plasma flow patterns. The EISCAT facility, with radars near Tromsø on mainland Norway and Longyearbyen on Svalbard, measures a number of plasma parameters, including electron density and electron and ion temperatures, while the SuperDARN radars give the line-of-sight velocity of the F-region convection flows. These flow observations are then inverted by a spherical harmonic fitting technique, described by Ruohoniemi and Baker (1998), to produce polar-projection maps of the high-latitude electric potential that are representative of the ionospheric convection patterns.

3.2.4   Summaries of Recent Case Studies

The main results from recent collaborative multi-instrument studies, led by the University of Wales Aberystwyth, are summarised here. The first two cases relate to the dayside and nightside sectors respectively when the plasma was drawn antisunward by the convective flow through the polar region. The third presents observations of the plasma distribution in the afternoon sector when IMF B_z was positive. It provides for the first time evidence of a tongue of enhanced plasma density circulating the polar cap, which is closed to the inflow of plasma under conditions of B_z northwards.

3.2.4.1 Dayside Polar Ionosphere: Source Region of the Tongue of Ionisation

A region of enhanced F-region ionospheric densities was observed by the EISCAT Svalbard radar (ESR) at high latitudes near local magnetic noon under conditions of IMF B_z negative (Pryse et al., 2004). The source of the plasma was investigated using optical, spacecraft, radar and tomographic instrumentation. In this instance it was possible to rule out *in situ* soft-particle precipitation as the production mechanism. The equivalent vertical total electron content (TEC) measured at the four sites of the UWA Scandinavian tomography chain (Figure 3-7) show a latitudinally-restricted region of enhanced ionisation with maximum at about 73ºN-74ºN, superimposed on the gradient of the dayside ionosphere created by photo-production. The study demonstrated that this region of increased ionisation at sub-auroral latitudes, seen during three consecutive satellite passes, was the likely source of the plasma enhancement observed by the ESR at polar latitudes, with the flux tubes being drawn poleward by the convective flow. The build-up of densities at the lower latitudes was proposed to occur at the equatorward edge of the afternoon high-latitude convection cell, where plasma is exposed to sunlight for an extended period as it flows slowly sunward toward magnetic noon under the influence of the combined effects of high-latitude convection and co-rotation.
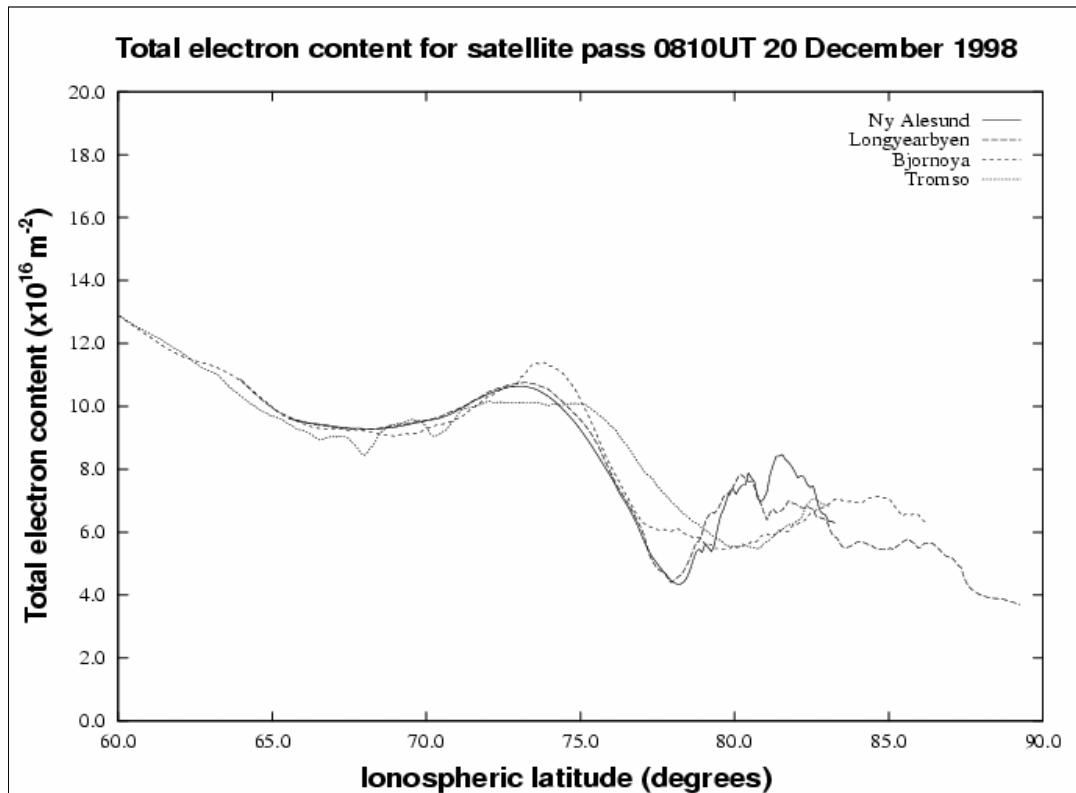
Figure 3-7: Equivalent Vertical Total Electron Content Measured at the Four Receiving Sites of the Scandinavian Tomography Chain During a NIMS Satellite Pass that Crossed Latitude 75ºN at 0810UT on 20 December 1998. (Pryse et al., 2004)

3.2.4.2 Nightside Ionosphere: Reconfiguration of a Polar Patch into a Boundary Blob

A long-lived, high-altitude, cold plasma enhancement observed in the anti-sunward convective flow across the polar cap was studied using measurements from the tomography experiment and the SuperDARN and EISCAT radar facilities (Pryse et al., 2006a). A projection of the feature to later times suggested that it was reconfigured in the Harang discontinuity to form an enhancement that was elongated in longitude in the evening sunward return flow of the high-latitude convection. This was supported by a second tomography image, some 45 min later, which showed a boundary blob almost coincident with the

projection. The tomographic reconstruction supporting the interpretation also showed another enhancement equatorward of the reconfigured blob, likely to have been produced by soft precipitation. Taken together, the observations (Figure 3-8) indicated that the two mechanisms proposed previously in the literature for the production of boundary blobs, were operating simultaneously to form two distinct density features separated slightly in latitude. The finding is of relevance to radio propagation applications as boundary blobs often form the poleward boundary of the main ionisation trough in this time sector.
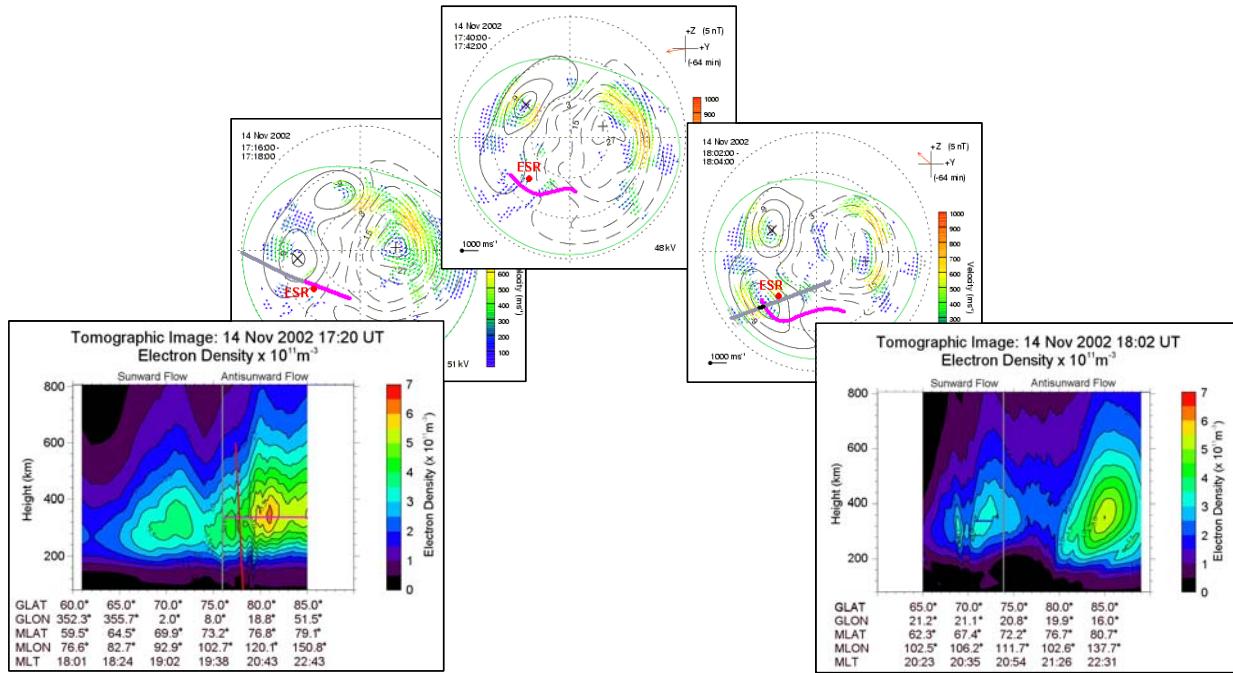
**Figure 3-8: A Polar Patch can be Seen in the Tomography Image from
a Satellite Pass at 1720UT on 14 November 2002 (left).**

The pink line marks its latitudinal extent which maps into the antisunward plasma flow shown on the SuperDARN electric potential patterns. Projection of the feature through successive SuperDARN observations (centre) shows that the line is reconfigured by the flow, becoming elongated in longitude in the sunward return flows of the dusk and dawn cells. A second tomography image at 1802UT (right) shows that the pink projection is essentially collocated with a boundary blob centred ~72ºN. A second distinct boundary blob is also seen in the image at slightly lower latitudes. (Pryse et al., 2006a)

3.2.4.3   A Tongue of Ionisation with IMF B_z Positive

A series of twelve consecutive satellite passes, under conditions of stable IMF $B_z>0$, from the tomography chain of the Applied Research Laboratories (ARL), University of Texas at Austin, in Greenland and the UWA chain in Scandinavia showed a persistent high-latitude density enhancement that was investigated by Middleton et al., 2005a. The images (samples of which are given in Figure 3-9) were interpreted as evidence for a tongue of ionisation when IMF B_z was positive. Even though the polar cap was closed to plasma inflow from lower latitudes, the dayside photoionisation drawn poleward by viscous convection cells was shown to flow around the periphery of the polar cap. This formed a band of enhanced densities that decreased with increasing time displacement from magnetic noon, due to chemical recombination.
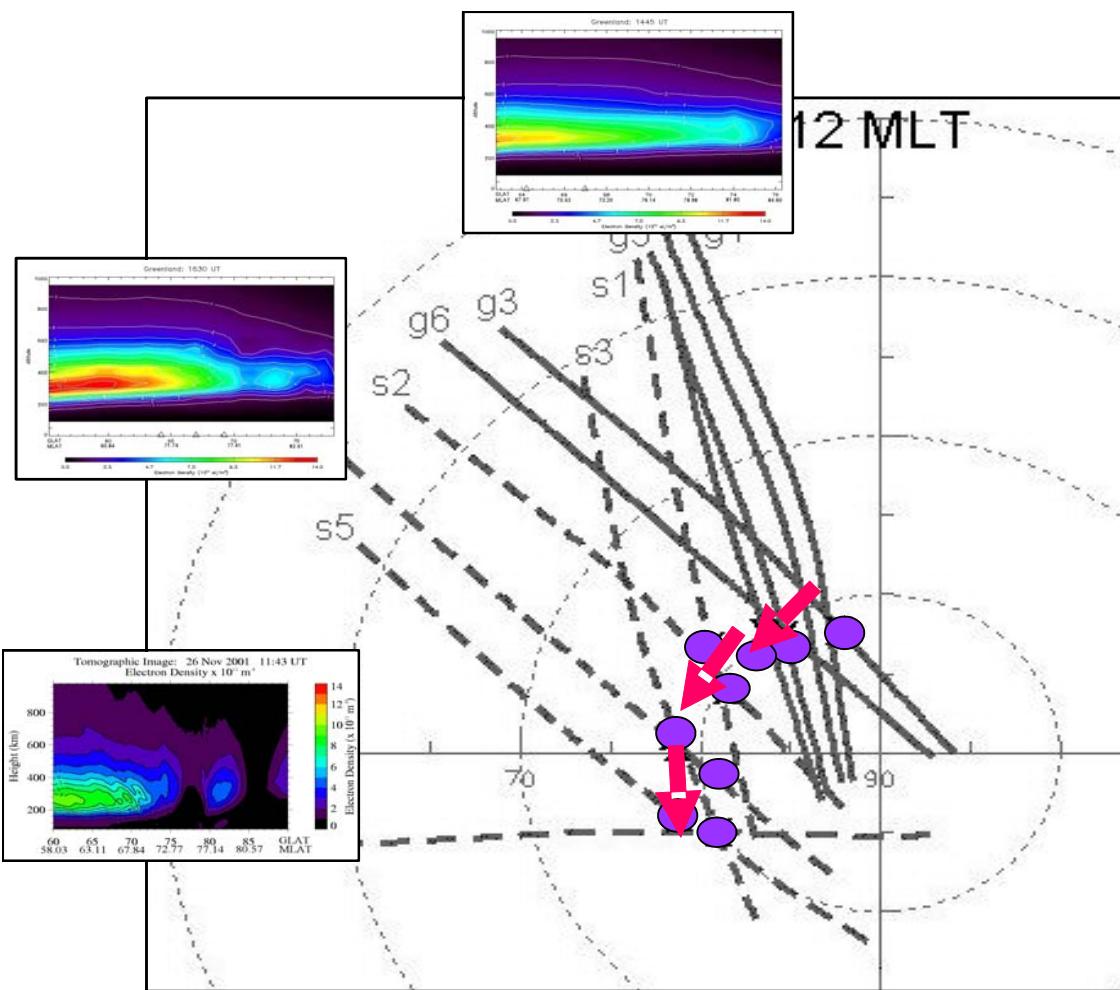
**Figure 3-9: Sample Tomography Images from the Greenland and Northern Scandinavian
Tomography Receiver Chains for 26 November 2001 Showing Cross-Sections
Through a Tongue of Ionisation Drawn around the Dusk-Side Periphery
of the Polar Cap under Conditions of Stable IMF $B_z > 0$.**

The purple dots on the satellite trajectories show the locations of the density enhancements observed in the series of satellite passes and the pink arrows show the antisunward plasma drift implied from DMSP horizontal drift measurements. (Middleton et al., 2005a)

3.2.5   Tomography and the CTIP Model: A Case Study

Physical models have an important role to play in aiding understanding of the relative contributions of the various physical processes underlying the complex plasma structure of the ionised atmosphere. The Coupled Thermosphere Ionosphere Plasmasphere (CTIP) model was developed over many years, by the University of Sheffield and University College London integrating a thermospheric model (Fuller-Rowell and Rees, 1980) with models of the plasmasphere and high-latitude ionospheres (Quegan et al., 1982).

A tomography reconstruction from a study relating observations and modelling (Middleton et al., 2005b) is shown in Figure 3-10. The image, for a satellite pass that crossed 75ºN latitude at 0007UT on 13 December 2001, clearly reveals a structured nighttime ionosphere with a large enhancement between 62°MLAT and 70.5°MLAT where the electron density exceeds $7 \times 10^{11} \text{m}^{-3}$. The darkened winter nighttime conditions at these latitudes rule out direct production by solar EUV as a source mechanism for the increased densities.
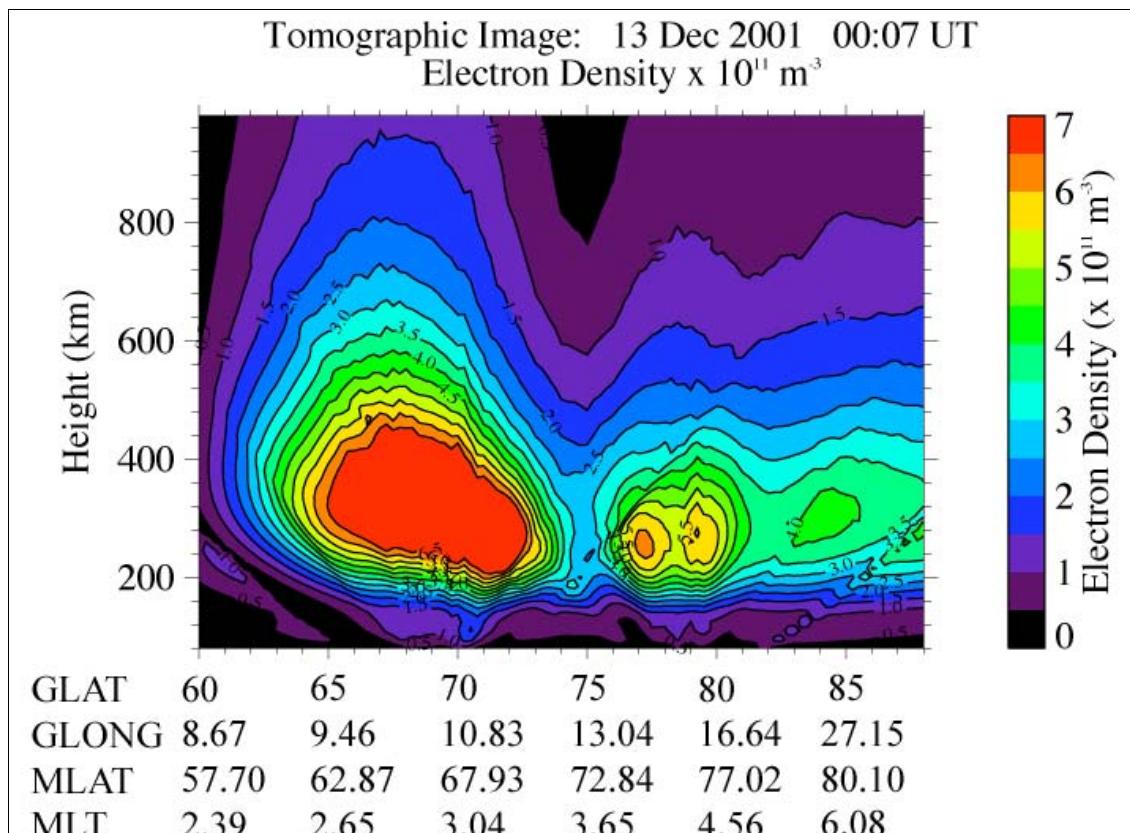
Figure 3-10: Tomographic Image for a Satellite Pass at 0007UT on 13 December 2001. A large density enhancement can be seen in the equatorward section of the field-of-view.

The feature was observed in five consecutive tomography images between 2305UT on 12 December and 0243UT on 13 December, three from the UWA tomography chain in Scandinavia and the other two by stations in the UK also operated at the time by UWA. The trajectories of the five satellite passes are shown in Figure 3-11, mapped on a SuperDARN electric potential pattern that was representative of the convection flow during the time of interest. The mapping of the maximum density in the enhancement for each image is indicated by a coloured dot. The dots lie mainly in the return flow of the dawn cell, indicating that the flux tubes are likely to have convected from the polar cap. Simultaneous observations by the EISCAT incoherent scatter radar on mainland Norway confirmed the high densities, but the corresponding cold electron temperatures ruled out in situ precipitation as the production mechanism. Understanding the origins of such enhancements is of interest, not only to the interpretation of the physical processes operating in the high-latitude ionosphere, but also to radio propagation applications, as this large ionisation feature forms the poleward wall of the main ionisation trough in the post-midnight sector.
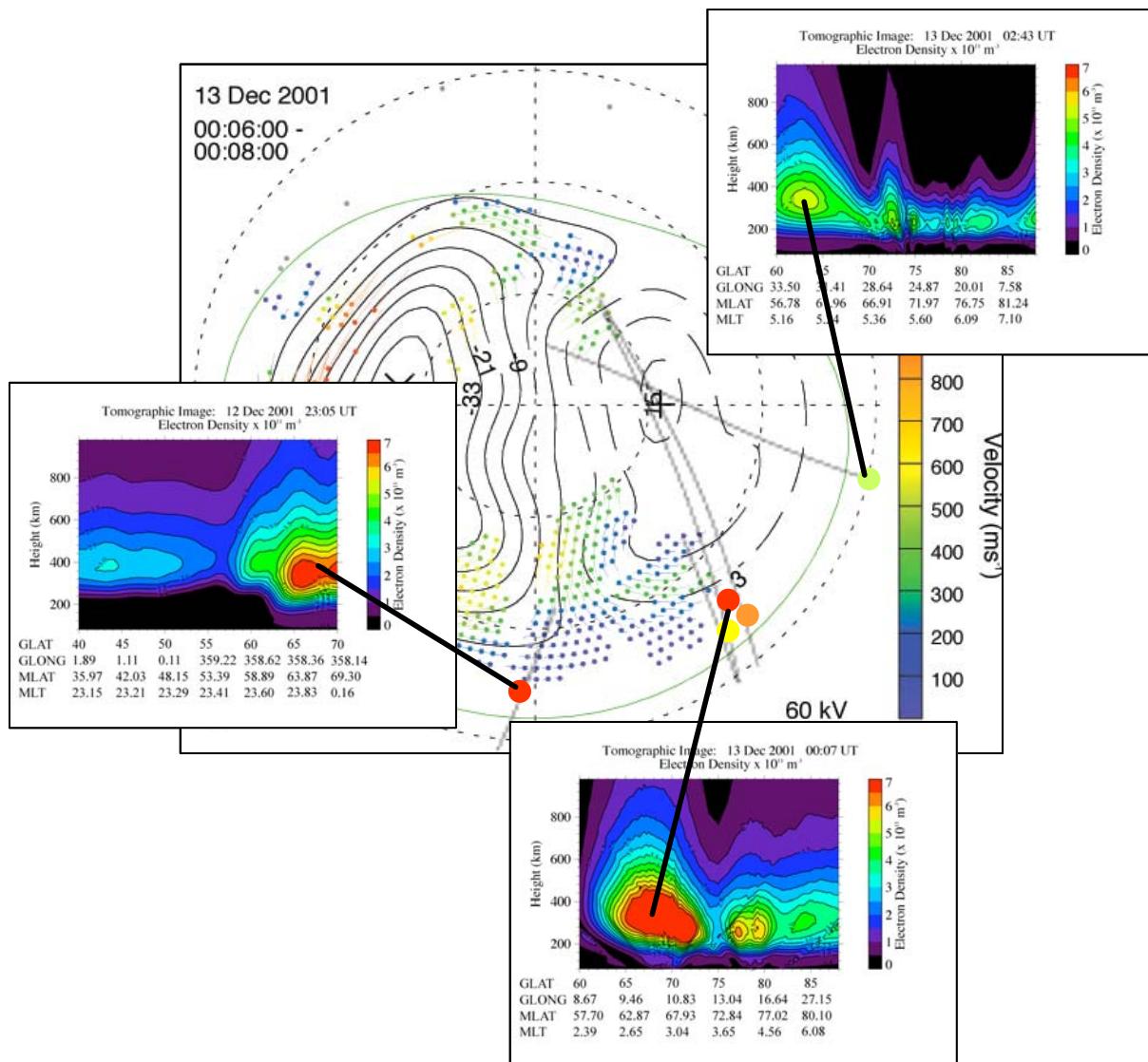
Figure 3-11: Sample Tomography Images between 2305UT on 12 December 2001 and 0243UT on 13 December 2001. The locations of the maximum electron density and mapped trajectories for all five satellite passes are shown superimposed on a representative electric potential (plasma flow) pattern measured by the SuperDARN radars.

The CTIP model provides support for the interpretation. Figure 3-12 shows a sequence of output plots of F_2-layer electron density, at 3-hourly intervals, from a model run with input parameters appropriate to the geophysical conditions at the time of observations. Magnetic noon is at the top of each panel, with magnetic midnight at the bottom, dawn is on the right-hand-side and dusk on the left. The latitude scale extends from 50ºMLAT on the outer circumference to the magnetic pole at the centre. The small white region near the centre of each panel indicates the location of the geographic pole. The white line on each panel shows the approximate location of the Scandinavian tomography receiver chain. Panels f) and g) for the UT intervals of relevance to the study show a TOI originating on the dayside and being drawn into the nightside, and then around into the dawn side at lower latitudes. It can be seen that at 0UT to 03UT the observations are clearly sampling a region of plasma drawn over the polar cap in a TOI.
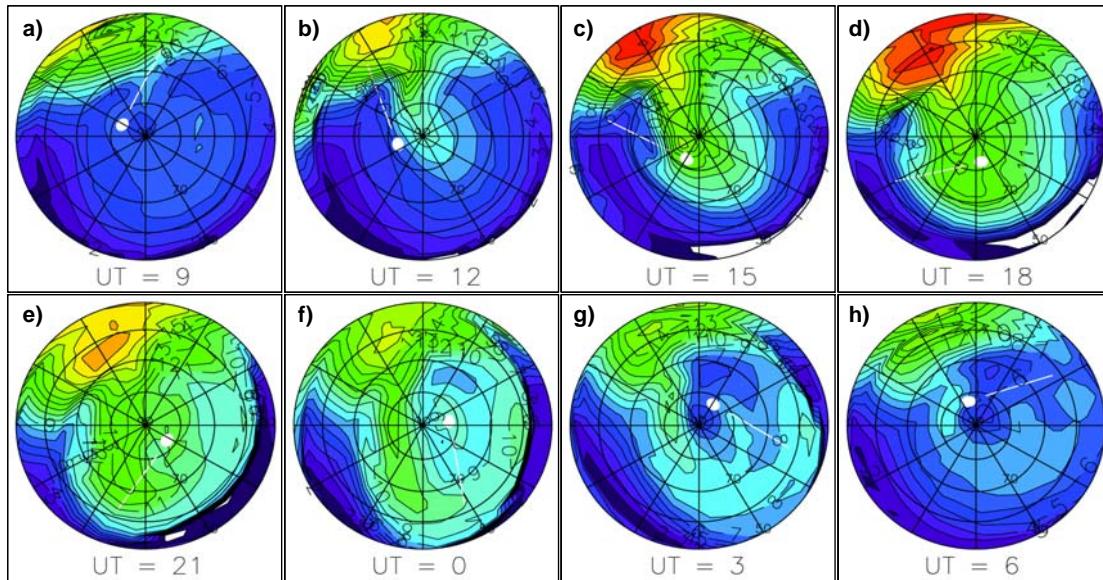
Figure 3-12: Electron Densities in the Polar F$_2$-Layer from the CTIP Model Plotted as a Function of MLT and MLAT. The latitude scale extends from 50º to the geomagnetic pole, geomagnetic noon is at the top of each circle and the panels are at 3-hourly intervals of UT.

The panels of Figure 3-12 throughout the day show that the TOI maximises near 18-21UT when Europe is entering the evening/nightside but the feature is much less marked around 06-09 UT when the European sector is on the dayside. This variation is a consequence of the offset between the geographic and geomagnetic reference frames; the photo-production is controlled by solar processes in the geographic frame while the high-latitude convection is driven in the geomagnetic frame. A larger proportion of the convection pattern is therefore in sunlight during the 18-21UT interval than in the earlier interval. The modelling thus suggests that the effect of the TOI in the evening sector is likely to be more prominent at European longitudes than in the US sector, whilst the converse is true for the 06-09UT interval. The results are consistent with previous work by Bowline et al. (1996) where three locations in different longitude sectors were considered, and which indicated that the TOI was expected to be prominent at Ny-Ålesund in the evening with B_y negative.

3.2.6 Statistical Analysis for Model Verification and Development

Physical models provide a valuable resource for interpretation of ionospheric observations, although measurements are also crucial for the verification and development of both physical, parameterised and empirical models. Preliminary steps have been made in the analysis of statistical results from tomographic observations for comparisons with models.

3.2.6.1 Dayside High-Latitude Trough

The dayside high-latitude trough is believed to be the extension of the main or mid-latitude nighttime trough into the earlier time sector of magnetic afternoon. It is a persistent feature of the wintertime auroral ionosphere at latitudes of about 70ºN to 75ºN. Radio tomography observations from December 2001 were used to map the location and latitudinal structure of this trough under quiet geomagnetic conditions (Kp≤2) near winter solstice (Pryse et al., 2005). The red curves and error bars respectively in Figure 3-13 show the median latitudinal distributions and associated inter-quartile ranges for N$_m$F$_2$. The CTIP model was run with input parameters representative of the observing conditions to examine the influences of both precipitation and convection on the model. The curves of Figure 3-13 correspond to four different combinations as follows: both precipitation energy input and high-latitude convection switched-on (DecCP), precipitation off

but convection on (DecC), precipitation on but convection off (DecP), and both precipitation and convection switched off (Dec).
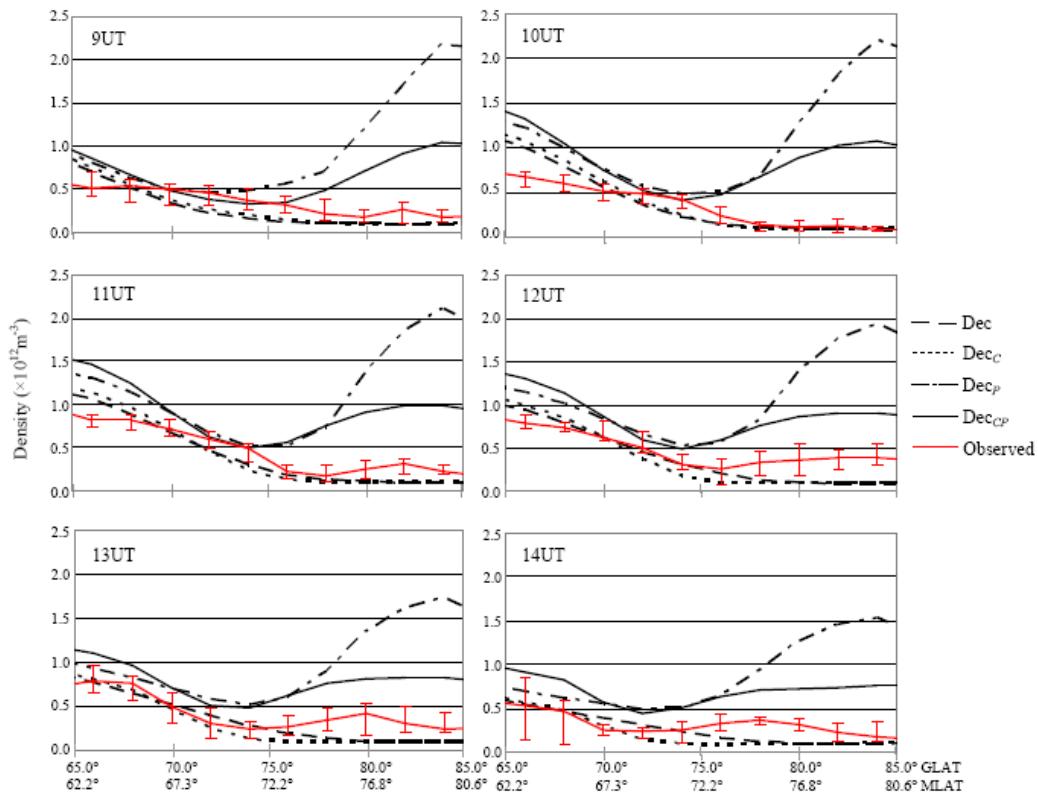
Figure 3-13: Latitudinal Variation of N_mF_2 in 1-hour Bins of UT Obtained from Tomography Observations in Scandinavia During December 2001 with Kp≤2. Corresponding outputs from the CTIP model are also shown with the explanations for the curves given in the text.

It can be seen that the model reproduces the general behaviour of the observed troughs, with photoionisation responsible for the equatorward wall, precipitation ionisation on the poleward wall, and a general decrease in trough latitude with increasing time. However, closer inspection reveals that the auroral ionisation is substantially overestimated in the model, the latitude of the trough minimum is generally displaced from that observed, and the equatorward migration of the trough with UT spans a greater latitudinal range in the observations than in the model. A further discrepancy arises from the shape of the equatorward trough wall, with the observed densities being maintained at a higher level than those predicted from photoionisation and a steeper decrease into the trough minimum.

3.2.6.2   Parameterisation of the Main Ionospheric Trough in the European Sector

Another investigation (Pryse et al., 2006b) was directed towards representing the latitudinal position and shape of the main or mid-latitude trough, observed during the evening and nighttime at UK latitudes, in such a way that the observations could be used for direct comparison and validation of ionospheric models. Satellite transmissions from the NIMS constellation were monitored at three tomography stations in the UK for a year between September 2002 and August 2003. The vertical total electron content as a function of latitude was determined for the passes by integration through the tomographic images. The TEC measurements were then used to characterise the structure and dynamics of the main ionospheric trough in terms of a set of defined parameters. The observations confirmed the well-established equatorwards movement of the trough throughout the night and with increased geomagnetic activity.

The TEC values in the trough minimum were generally consistent at about 2 to 3 TECU, while the depth of the trough was much more dependent on Kp in the evening hours than after midnight. The trough half-width on the equatorward side of the minimum showed greater variability than the corresponding parameter on the poleward side. Figure 3-14 gives a sample of the results showing the median trough profiles plotted from the defined parameters for conditions appropriate to a mid-range of geomagnetic activity, with Kp between 3- and 4-. Corresponding plots were also obtained for low and high Kp (Pryse et al., 2006b). It can be seen from Figure 3-14 that the trough retains a notch-like shape throughout most of the night for geomagnetic activity in the mid-Kp range. The gradient of the equatorward wall becomes less steep at later times with the ionisation at the mid-latitudes decreasing during the night. The poleward wall is often steeper than that on the equatorward side of the minimum and is usually surmounted by a boundary blob maximum. The parameters used in this study have been defined in such a way that the experimental results from the study can be compared directly with the output of models. Validation of empirical and parameterised models of the ionosphere, developed for the mitigation of propagation effects in radio systems applications is of importance in the vicinity of the main trough. Few current models attempt to replicate the trough and, for those that do, getting the trough in the wrong place or with the wrong shape may simply confound the problem.
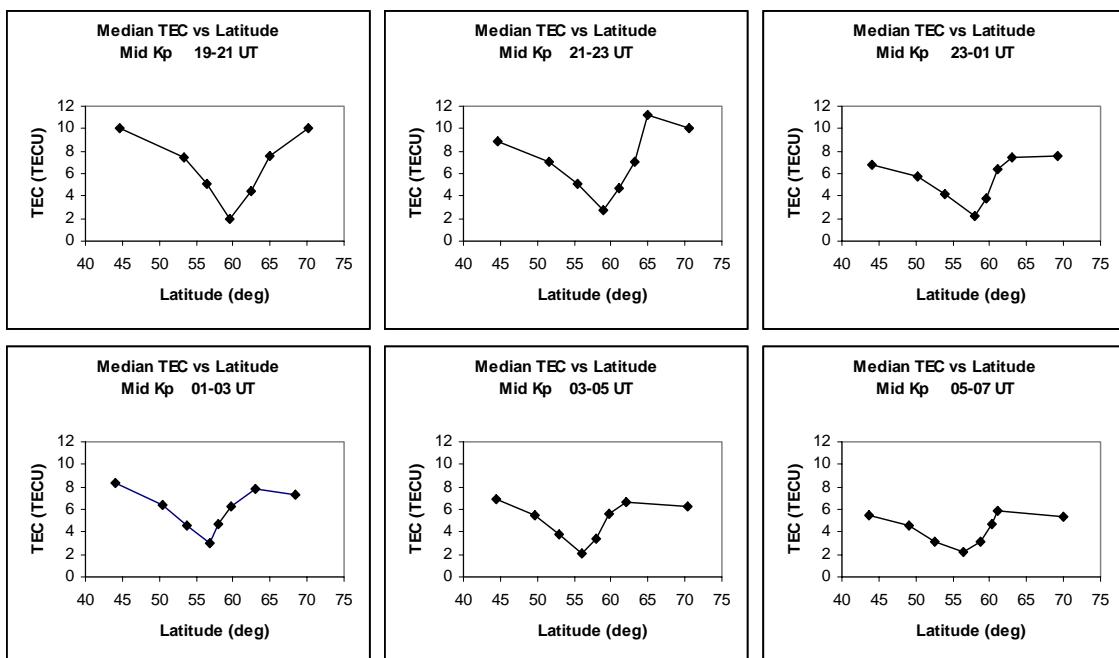
Figure 3-14: Median TEC versus Median Latitude of the Trough Parameters for Mid-Range Kp, Obtained from Tomographic Observations in the UK between September 2002 and August 2003 (Pryse et al., 2006b).

3.2.7   Summary and Open Issues

Results have been summarised here from collaborative studies led by the University of Wales Aberystwyth into the spatial distribution of the ionospheric plasma at high latitudes. Observational evidence has been discussed in support of dayside ionisation being drawn by the convective flow into the polar cap. It was shown that with IMF B_z positive the dayside ionisation can be drawn around the periphery of a polar cap that is closed to the inflow of plasma. In addition, observations and modelling studies have demonstrated that polar cap plasma may be reconfigured on the nightside to form the poleward wall of the main ionisation trough. The ongoing experimental programme forms a basis for longer-term statistical studies of the polar and auroral ionospheres, of relevance to the development and

validation of physical and parameterised models of the regions. Definitive studies have been made in this respect on the form and location of the latitudinal structure of the main trough.

With the International Polar Year (IPY) imminent, it is now timely to consider the studies within a broader framework. The International Ionospheric Tomography Community has submitted an Expression of Intent that now forms part of the "Heliosphere impact on geospace" programme. The radio tomography programme comprises receiver chains in Alaska, Greenland, Scandinavia (polar and auroral) and Russia. A preliminary study reported by Kersley et al. (2005), using observations from the receiver chains in Alaska, Greenland and Scandinavia, illustrated the potential of multi-chain tomography for near-simultaneous observations in different longitude sectors opening the way for longer term studies.

Several specific issues that remain to be addressed include:

- *Origin of enhanced plasma on the dayside*

Where does the plasma build-up? What is the role of the combined effects of the high-latitude convection and co-rotation flow in retaining plasma in sunlight for an extended time period?

- *Mechanisms of entrainment of plasma into the high-latitude convection pattern*

Does this form a continuous tongue of ionisation? Alternatively, is it influenced by magnetic reconnection that creates a succession of patches at sub-polar latitudes as proposed by Moen et al. (2006)?

- *Effect of nightside reconnection*

What is the effect of the closure of the geomagnetic field on the structure of the plasma flowing out of the polar cap on the nightside?

- *Offset of the geomagnetic and geographic reference frame*

Do observational data support plasma of dayside origin being more prominent at nighttime in the European sector than the US sector? Is this the reason for most of the large SED events reported to date, of concern to GPS applications, originating in the US sector?

- *Influence of plasma flow patterns under conditions of stable IMF B_z positive*

How does the flow pattern affect the spatial distribution of plasma at high latitudes? Can the flow pattern and the plasma distribution be modelled?

- *Main ionisation trough*

What are the mechanisms that form and sustain the poleward wall of the trough on the nightside? Are these the same in different longitude sectors?

- *Particle precipitation*

How does *in situ* electron density production by particle precipitation modulate the spatial distribution of the ionospheric plasma at auroral and polar latitudes?

3.2.8   Acknowledgements

for the studies has come from the UK Particle Physics and Astronomy Research Council, the UK Radiocommunications Agency and NorthWest Research Associates Inc., USA.

3.2.9   References

Anderson, D.N., Buchau, J., and Heelis, R.A.: Origin of density enhancements in the winter polar cap ionosphere, Radio Sci., 23, 513-519, 1988.

Bowline, M.D., Sojka, J.J., and Schunk, R.W.: Relationship of theoretical patch climatology to polar cap patch observations, Radio Sci., 31, 635-644, 1996.

Buchau, J., Weber, E.J., Anderson, D.N., Carlson, Jr., H.C., Moore, J.G., Reinisch, B.W., and Livingston, R.C.: Ionospheric structures in the polar cap: their origin in relation to 250-MHz scintillation, Radio Sci., 20, 325-338, 1985.

Crowley, G.: Critical review of ionospheric patches and blobs, URSI Review of Radio Science 1993-1996, edited by: Stone, W. R., 619-648, 1996.

Cowley, S.W.H.: Excitation of flow in the Earth's magnetosphere-ionosphere system: observations by incoherent scatter radar, in Polar Cap Boundary Phenomena, eds. J. Moen et al., NATO Advanced Study Institute Series, Kluwer Academic Press, Dordrecht, The Netherlands, 509, 127-140, 1998.

Dungey, J.W.: Interplanetary magnetic field and auroral zones, Phys. Rev. Lett., 6 (2), 47-48, 1961.

Fuller-Rowell, T.J. and Rees, D.: A three-dimensional, time-dependent, global model of the thermosphere, J. Atmos. Sci., 37, 2545-2567, 1980.

Greenwald, R.A., Baker, K.B., Dudeney, J.R., Pinnock, M., Jones, T.B., Thomas, E.C., Villain, J.-P., Cerisier, J.-C., Senior, C., Hanuise, C., Hunsucker, R.D., Sofko, G., Koehler, J., Nielsen, E., Pellinen, R., Walker, A.D.M., Sato, N., and Yamagishi, H.: Darn/Superdarn: A global view of the dynamics of high-latitude convection, Space Sci. Rev., 71, 761-796, 1995.

Jones, D,G., Walker, I.K., and Kersley, L.: Structure of the poleward wall of the trough and the inclination of the geomagnetic field above the EISCAT radar, Ann. Geophysicae, 15, 740-746, 1997.

Kersley, L., Pryse, S.E., Denton, M.H., Bust, G., Fremouw, E., Secan, J., Jakowski, N., and Bailey, G.J.: Radio tomographic imaging of the northern high-latitude ionosphere on a wide geographic scale, Radio Sci., 40, RS5003, doi: 10.1029/2004RS003103, 2005.

Lockwood, M.: Identifying the open-closed field line boundary, in Polar Cap Boundary Phenomena, eds. J. Moen et al., NATO Advanced Study Institute Series, Kluwer Academic Press, Dordrecht, The Netherlands, 509, 73-90, 1998.

McEwen, D.J., and Harris, D.P.: Occurrence patterns of F layer patches over the north magnetic pole, Radio Sci., 31, 619-628, 1996.

Middleton, H.R., Pryse, S.E., Dewis, K.L., Wood, A.G., and Balthazor, R.: Signatures of space weather processes in the northern polar ionosphere: Radio Tomography and the CTIP model, 2nd European Space Weather Week: ESWW-II, ESA-ESTEC, Noordwijk, The Netherlands, November 2005a (poster).

Middleton, H.R., Pryse, S.E., Kersley, L., Bust, G.S., Fremouw, E.J., Secan, J.A., and Denig, W.F.: Evidence for the tongue of ionisation under northward IMF conditions, J. Geophys. Res., 110(A7), A07301 10.1029/2004JA010800, 2005b.

Moen, J., Carlson, H.C., Oksavik, K., Nielsen, C.P., Pryse, S.E., Middleton, H.R., McCrea, I.W., and Gallop, P.: EISCAT observations of plasma patches at sub-auroral cusp latitudes, Ann. Geophysicae, 24(9), 2363-2374, 2006.

Pryse, S.E.: Radio tomography: A new experimental technique, Surveys in Geophysics, 24, 1-38, 2003.

Pryse, S.E., Sims, R.W., Moen, J., Kersley, L., Lorentzen, D., and Denig, W.F.: Evidence for solar-production as a source of polar-cap plasma, Ann. Geophsyicae, 22, 1093-1102, 2004.

Pryse, S.E., Dewis, K.L., Balthazor, R.L., Middleton, H.R., and Denton, M.H.: The dayside high-latitude trough under quiet geomagnetic conditions: Radio tomography and the CTIP model, Ann. Geophysicae, 23(4), 1199-1206, 2005.

Pryse, S.E., Wood, A.G., Middleton, H.R., McCrea, I.W., and Lester, M.: Reconfiguration of polar-cap plasma in the magnetic midnight sector, Ann. Geophysicae, 24(8), 2201-2208, 2006a.

Pryse, S.E., Kersley, L., Malan, D., and Bishop, G.J.: Parameterization of the main ionospheric trough in the European sector, Radio Sci., 41, RS5S14, doi:10.1029/2005RS003364, 2006b.

Quegan, S., Bailey, G.J., Moffett, R.J., Heelis, R.A., Fuller-Rowell, T.J., Rees, D. and Sprio, R.W.: A theoretical study of the distribution of ionisation in the high-latitude ionosphere and the plasmasphere: first results on the mid-latitude trough and the light-ion trough, J. Atmos. Terr. Phys., 44, 619-640, 1982.

Reiff, P.H., and J.L. Burch, IMF By-dependent plasma flow and Birkeland currents in the dayside magnetosphere: 2. A global model for northward and southward IMF, J. Geophys. Res., 90 (NA2), 1595-1609, 1985.

Rishbeth, H., and Williams, P.J.S.: The EISCAT ionospheric radar: the system and early results, Q. J. R. Astr. Soc., 26, 478-512, 1985.

Robinson, R.M., Tsunoda, R.T., and Vickrey, J.F.: Sources of F-region ionisation enhancements in the nighttime auroral zone, J. Geophys. Res., 90, 7533-7546, 1985.

Rodger, A.S., Pinnock, M., Dudeney, J.R., Baker, K.B., and Greenwald, R.A.: A new mechanism for polar patch formation, J. Geophys. Res., 99, 6425-6436, 1994.

Ruohoniemi, J.M., and Baker, K.B.: Large-scale imaging of high-latitude convection with Super Dual Auroral radar Network HF radar observations, J. Geophys. Res., 103, 20797-20806, 1998.

Sims, R.W., Pryse, S.E., and Denig, W.F.: Spatial structure of summertime ionospheric plasma near magnetic noon, Ann. Geophysicae, 23, 25-37, 2005.

Sojka, J.J., Bowline, M.D., Schunk, R.W., Decker, D.T., Valladares, C.E., Sheehan, R., Anderson, D.N., and Heelis, R.A.: Modelling polar-cap F-region patches using time-varying convection, Geophys. Res. Lett., 20, 1783-1786, 1993.

Valladares, C.E., Basu, S., Buchau, J., and Friis-Christensen, E.: Experimental evidence for the formation and entry of patches into the polar cap, Radio Sci., 29, 167-194, 1994.

Weber, E.J., Buchau, J., Moore, J.G., Sharber, J.R., Livingston, R.C., Winningham, J.D., and Reinisch, B.W.: F-layer ionization patches in the polar cap, J. Geophys. Res., 89, 1683-1694, 1984.

3.3 THE ELECTRON DENSITY ASSIMILATIVE MODEL (EDAM)

(M. J. Angling, QinetiQ, St. Andrew's Road, Malvern, Worcs, WR14 3PS, UK.)

3.3.1   Introduction

The last 15 years have seen a rapid growth in the development and use of ionospheric imaging systems. Traditional tomographic imaging techniques represent the ionosphere by means of a grid of pixels. Then, in the case of an integrated measurement, such as GPS TEC, the ith observation can be modelled by the sum of the electron density in the jth pixel (x_j) multiplied by the ray length within the jth pixel (H_{ij}) (Figure 3-15):

$$y_i = \sum_{j=1}^{n} H_{ij} x_j \tag{1}$$

or, in matrix notation:

$$\mathbf{y} = \mathbf{Hx} \tag{2}$$

where $\mathbf{H}$ is known as the observation operator, $\mathbf{x}$ is the rasterised representation of the ionosphere and $\mathbf{y}$ is the TEC.
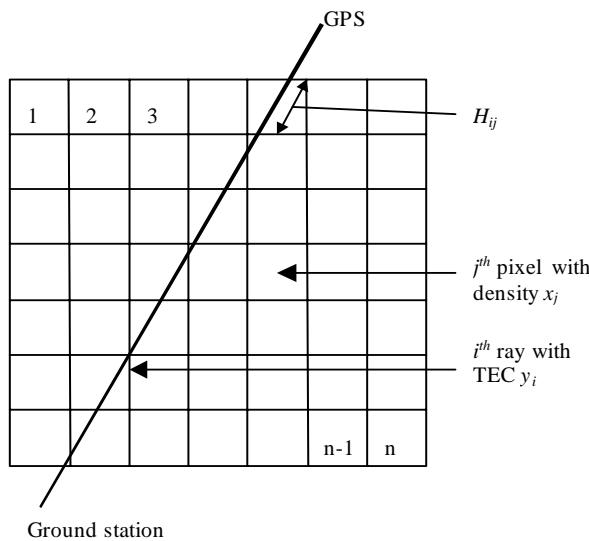
Figure 3-15: Diagram Illustrating the Construction of the Observation Operator that Relates a Slant TEC Measurement to a Pixel Representation of the Ionosphere.

It can be seen that this formulation of the imaging problem is entirely general and can, therefore, be easily extended to a 3D grid of voxels. Furthermore, there is no requirement that the slant TEC must be measured by a ground station, and consequently it is also simple to directly include measurements of TEC made in low earth orbit. The disadvantage of moving from 2D to 3D imaging is that a very large amount of data is required to be able to operate the traditional tomographical methods. In practice, and especially for GPS data (where the satellites are slow moving), the required quantity of data is not available, and other, more sophisticated, methods must be used. These methods must use *a priori* data to constrain the inversion from slant TEC to the electron density image. This *a priori* data may be a model of the ionosphere, or may take the form of assumptions about how the ionosphere can be decomposed into a set of basis functions.

Ideally, a 3D imaging system will use data provided by a range of ionospheric measurement techniques. Of particular note is data from the International GNSS Service (IGS) receiver network (Beutler et al., 1999). Many IGS stations provide data in hourly files with low latency (~90 minutes), which can be used to calculate slant total electron content (TEC). Radio occultation (RO) methods are also being increasingly investigated. RO measurements are made by monitoring transmissions from GPS satellites using receivers on Low Earth Orbiting (LEO) satellites and provide the potential of measuring refractivity profiles in regions where ground based sensors cannot easily be located, such as deep sea waters. The recent launch of RO instruments on the Constellation Observing System for Meteorology, Ionosphere & Climate (COSMIC) six satellite constellation has the potential to provide a large amount of ionospheric data (Hajj et al., 2000); again this will be available with a low latency (115 minutes). For both ground and space based data, data assimilation provides an optimal way of combining slant TEC measurements with an ionospheric model to provide a full 3D representation of the ionospheric electron density.

3.3.2   The Electron Density Assimilative Model

The Electron Density Assimilative Model (EDAM) has been developed (Angling and Cannon, 2004; Angling and Khattatov, 2006) to assimilate measurements into a background ionospheric model. This model is provided by PIM (Daniell et al., 1995) and the majority of the input data is GPS TEC derived from IGS stations. The assimilation is based on a weighted, damped least mean squares estimation. This is a form of minimum variance optimal estimation (also referred to as best linear unbiased estimation, BLUE) that provides an expression for an updated estimation of the state (known as the analysis) that is dependent upon an initial estimate of the state (the background model), and the differences between the background model and the observations (Menke, 1989). The error covariance matrices of the background model and the observations are also included to control the relative contributions of the background and the observations to the analysis:

$$\mathbf{x_a} = \mathbf{x_b} + \mathbf{K}(\mathbf{y} - \mathbf{Hx_b})$$

(3)

$$\mathbf{K} = \mathbf{BH^T}\left(\mathbf{HBH^T} + \mathbf{R}\right)^{-1}$$

(4)

where $\mathbf{x_a}$ is the analysis, $\mathbf{x_b}$ is the background model, $\mathbf{K}$ is the weight matrix, $\mathbf{y}$ is the observation vector, $\mathbf{B}$ is the background error covariance matrix, and $\mathbf{R}$ is the error covariance matrix of the observations (Rodgers, 2000). $\mathbf{H}$ is the observation operator that relates the measurements to the state:

$$\mathbf{y} = \mathbf{Hx} + \boldsymbol{\varepsilon}$$

(5)

where $\boldsymbol{\varepsilon}$ is the observation error. The analysis error covariance matrix ($\mathbf{S}$) may also be calculated thus:

$$\mathbf{S} = \mathbf{B} - \mathbf{BH^T}\left(\mathbf{HBH^T} + \mathbf{R}\right)^{-1}\mathbf{HB}$$
$$\mathbf{S} = \mathbf{B} - \mathbf{KHB}$$

(6)

The assimilation is conducted using a magnetic coordinate system that remains fixed in space with respect to the sun. An assimilation time step of 15 minutes has been used and the electron density differences between the voxels of the analysis and the background model are propagated from one time step to the next by assuming persistence combined with an exponential decay. The time constant for this decay is set at 4 hours. Thus if the data feed is interrupted, the analysis will decay back to the background model.

A typical set of input data sites is shown in Figure 3-16. IGS stations are represented by dots, whilst USAF ionosondes are shown as triangles.
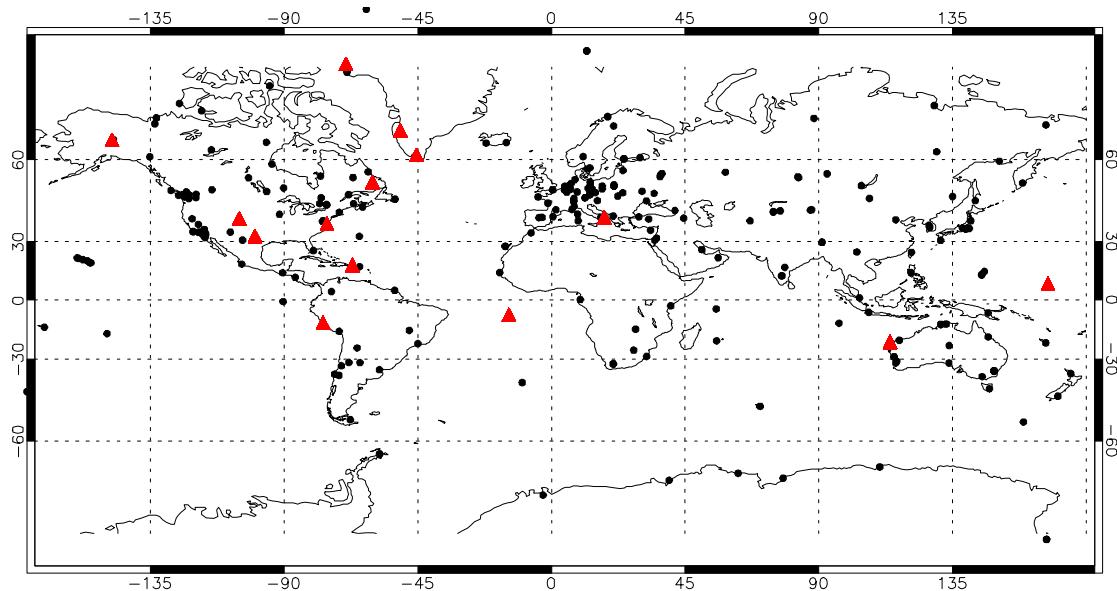
Figure 3-16: Map Showing Locations of IGS Stations (dots) and Vertical Ionosondes (triangles).

3.3.3   Testing

Testing has shown that, whilst the RMS errors in f_oF_2 may be reduced by the assimilative model (Figure 3-17), it is difficult to modify the vertical structure of the electron density grid using ground based TEC data alone (Figure 3-18) – this is a problem common to all assimilative approaches (Angling and Khattatov, 2006). It is anticipated that the introduction of RO data will provide much better vertical information and, therefore, improve the vertical representation of the ionosphere.
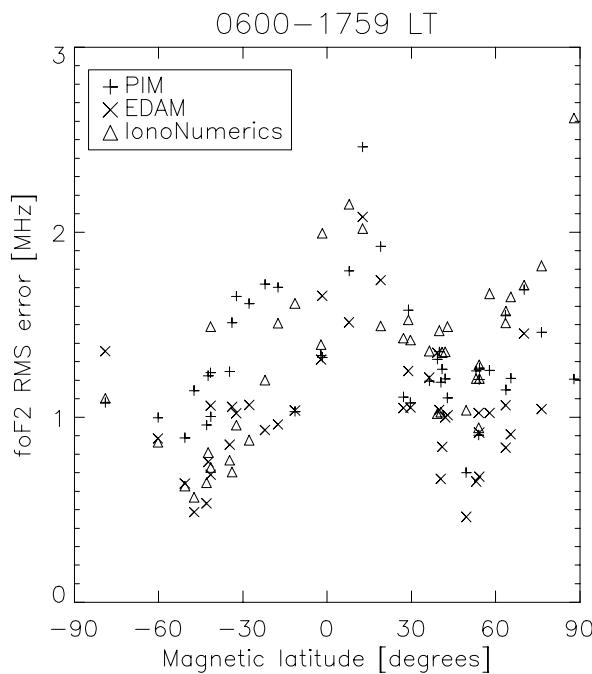
Figure 3-17: Daytime f_oF_2 RMS Error as a Function of Magnetic Latitude – from Angling and Khattatov, 2006.
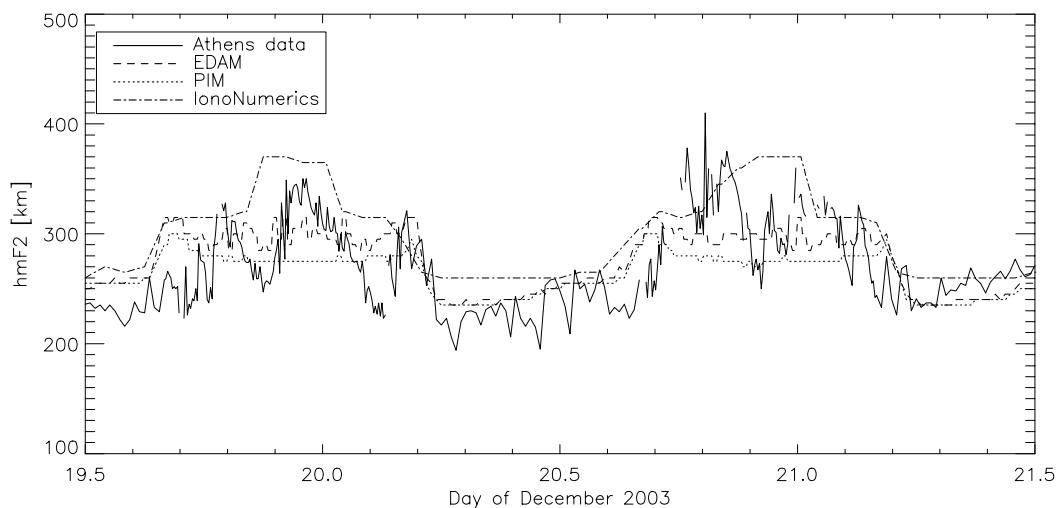
Figure 3-18: Example h$_m$F$_2$ Results for the Athens Ionosonde – from Angling and Khattatov, 2006.

Further testing has been undertaken using both ground based IGS data and space based data from COSMIC. For moderate (4/9/2006, Figure 3-19) and disturbed conditions (19-20/8/2006, Figure 3-20), EDAM generally performs better than PIM. As expected, the assimilation of the RO slant TEC data has the greatest impact on the assimilation results, though the improved RMS exhibited by EDAM ingesting just ground based IGS data does demonstrate that it is possible to reduce errors in the vertical structure of the ionospheric model using this data.
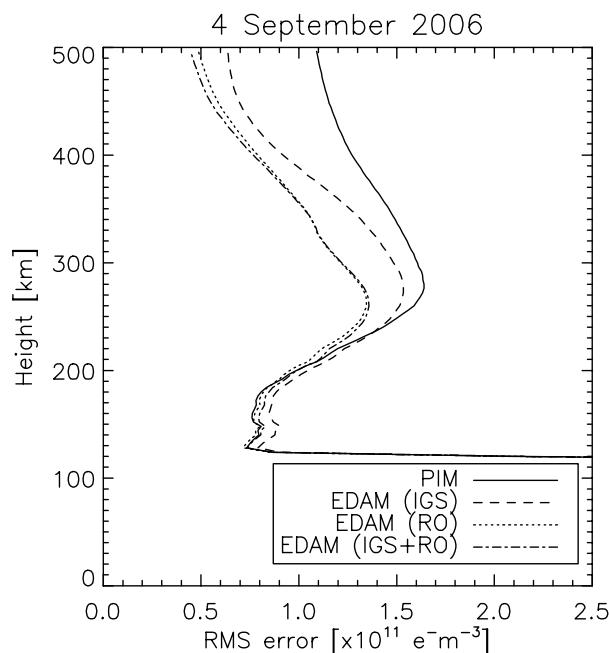
Figure 3-19: RMS Error between Ionospheric Models and Abel Transform Vertical Profiles at 5 km Height Steps on 4 September 2006. The models are: PIM (solid line); EDAM assimilating IGS data only (dashed line); EDAM assimilating COSMIC RO data only (dotted line); and EDAM assimilating IGS and COSMIC RO data (dot-dash line) – from Angling, 2007.
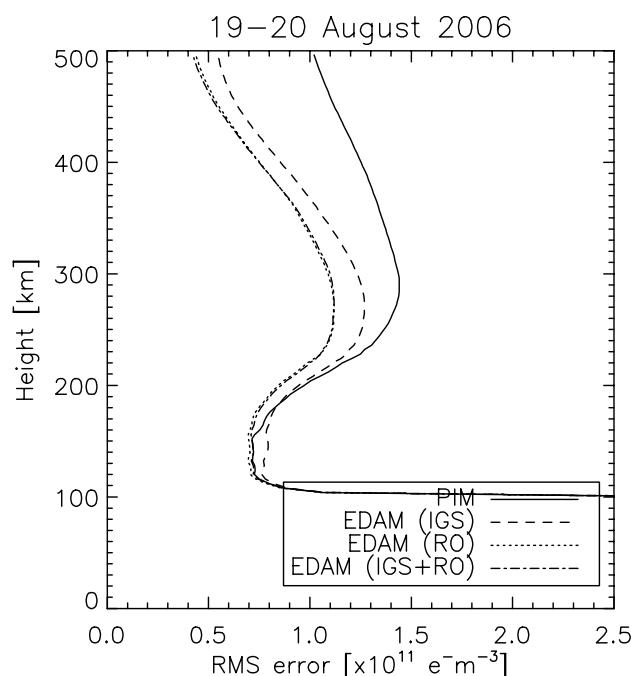
Figure 3-20: RMS Error between Ionospheric Models and Abel Transform Vertical Profiles at 5 km Height Steps on 19-20 August 2006. The models are: PIM (solid line); EDAM assimilating IGS data only (dashed line); EDAM assimilating COSMIC RO data only (dotted line); and EDAM assimilating IGS and COSMIC RO data (dot-dash line) – from Angling, 2007.

3.3.4   Conclusions

EDAM has been developed to provide an effective means for assimilating a range of difference ionospheric data types. Testing has demonstrated that EDAM can provide a better representation of the ionosphere than median models and that both ground and space based data are capable of improving the representation of the vertical structure of the ionosphere.

A number of other data assimilation models are in development or use. It would be greatly beneficial to have agreed data sets (i.e. of manually scaled ionograms) available to the community to facilitate comparative testing of these assimilative models.

3.3.5   Acknowledgements

EDAM has been developed under funding from the Operating Environment domain of the United Kingdom Ministry of Defence Corporate Science and Technology programme. IGS data was obtained from the IGS JPL and SOPAC Data Centres. Differential code biases were obtained from the Centre for Orbit Determination in Europe. Vertical ionosonde data was provided by the Air Force Research Laboratory and COSMIC data was provided by the COSMIC Data Analysis and Archive Centre.

3.3.6   References

Angling, M.J., First assimilations of COSMIC radio occultation data into the Electron Density Assimilative Model (EDAM), Submitted to Annales Geophysicae, 2007.

Angling, M.J., and P.S. Cannon, Assimilation of radio occultation measurements into background ionospheric models, Radio Science, 39 (RS1S08, doi: 10.1029/2002RS002819), 2004.

Angling, M.J. and B. Khattatov, Comparative study of two assimilative models of the ionosphere, Radio Science, 41, (RS5S20, doi: 10.1029/2005RS003372), 2006.

Belehaki, A. and L. Kersley, Statistical validation of a technique for estimating total electron content from bottomside ionospheric profiles, Radio Science, 41, (RS5003, doi: 10.1029/2005RS003433), 2006.

Beutler, G., M. Rothacher, S. Schaer, T.A. Springer, J. Kouba, and R.E. Neilan, The International GPS Service (IGS): An Interdisciplinary Service in Support of Earth Sciences, Adv. Space Res. Vol. 23, No. 4, pp. 631-635, 1999.

Daniell, R. E. Jr., L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka, R. W. Schunk, Parameterized ionospheric model: A global ionospheric parameterization based on first principles models, Radio Sci., 30(5), 1499-1510, 10.1029/95RS01826, 1995.

Hajj, G.A., L.C. Lee, X. Pi, L.J. Romans, W.S. Schreiner. P.R. Straus and C. Wang, COSMIC GPS Ionospheric Sensing and Space Weather, Special Issue of Terrestrial Atmos. Oceanic Sci., 11, 235-272, 2000.

Menke, W., Geophysical data analysis: discrete inverse theory, International Geophysics Series, San Diego, Academic Press, 1989.

Reinisch, B. W. and Huang, X.: Deducing topside profiles and total electron content from bottomside ionograms, Adv. Space Res., 27, 1, 23-30, 2001.

B.W. Reinisch, X. Huang, A. Belehaki, and R. Ilma, Using scale heights derived from bottomside ionograms for modelling the IRI topside profile, Advances in Radio Science, 2, 293-297, 2004.

Rodgers, C. D., Inverse methods for atmospheric sounding: theory and practice, Series on Atmospheric, Oceanic and Planetary Physics – Vol 2, World Scientific Publishing, 2000.

3.4 GPS SOUNDING OF THE IONOSPHERE ONBOARD CHAMP

(N. Jakowski, C. Mayer, V. Wilken, Deutsches Zentrum für Luft- und Raumfahrt (DLR) / Institut für Kommunikation und Navigation, Kalkhorstweg 53, Neustrelitz, Germany)

Abstract

Space based GPS measurements onboard Low Earth Orbiting (LEO) satellites provide a unique possibility for exploring the ionosphere on a global scale. Both the radio occultation measurements in the limb sounding mode and the navigation measurements using a zenith viewing GPS antenna provide the Total Electron Content (TEC) along numerous ray paths. TEC may effectively be used for reconstructing the spatial and temporal distribution of the electron density in the ionosphere and plasmasphere.

Reported are results obtained from radio occultation measurements on CHAMP which have provided more than 200,000 vertical electron density profiles so far. These observations contribute to a better understanding of the regular behaviour of the global ionosphere. Furthermore, the radio occultation measurements indicate irregular and/or wavelike structures in the ionosphere which may have severe impact on the functionality of radio systems.

A three-dimensional imaging of the electron density distribution near the CHAMP orbit plane between CHAMP and GPS orbit height is performed by using link related TEC data derived from dual frequency

navigation measurements onboard CHAMP. This type of measurements provides a good measure of the interaction of the solar wind with the global Earth's atmosphere, thus providing a good opportunity for studying this interaction via the magnetosphere.

Key words: GPS, Ionosphere, Plasmasphere, Radio Occultation, CHAMP.

3.4.1   Introduction

Innovative GPS based sounding techniques of the ionosphere have been established during the last decade. GPS receiver installed onboard Low Earth Orbiting (LEO) satellites have the capability to probe the ionosphere in the limb sounding mode (GPS radio occultation) and by using the navigation data for sounding the topside ionosphere / plasmasphere (see Figure 3-21).
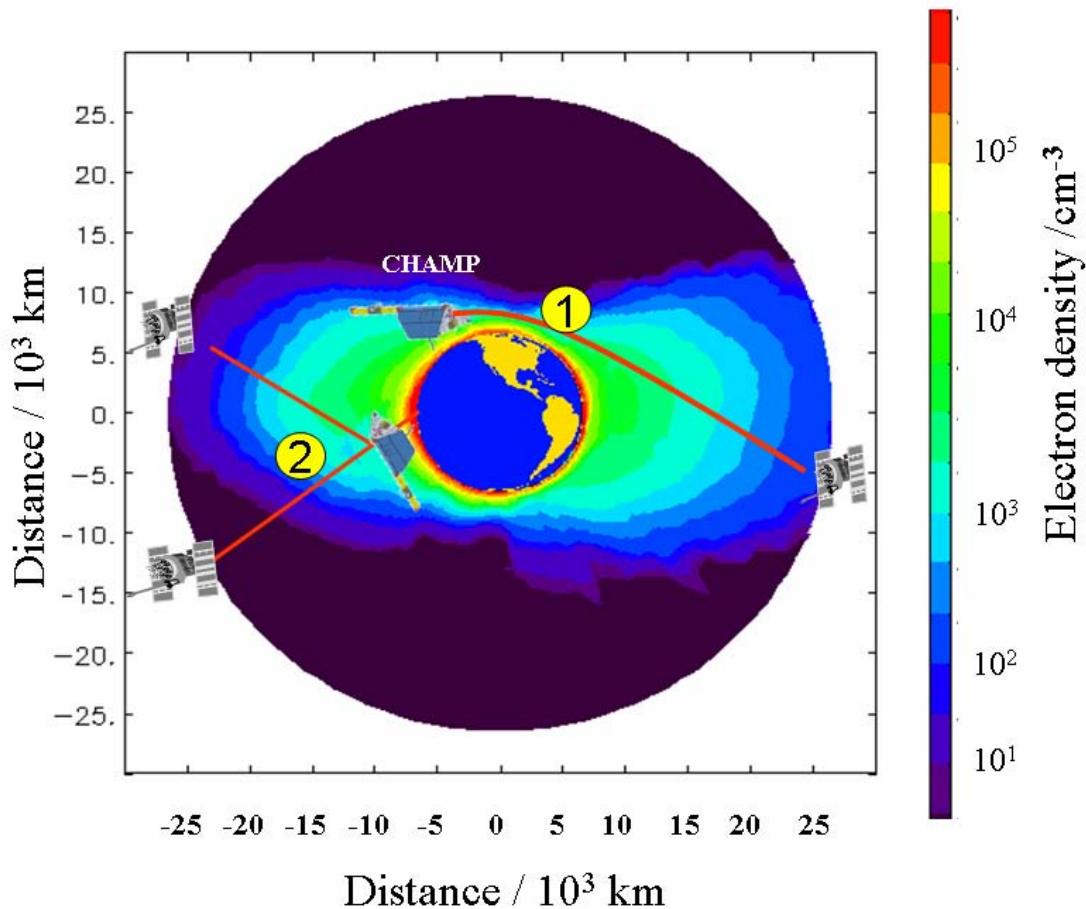
Figure 3-21: Illustration of GPS Measurement Techniques Used Onboard CHAMP for Sounding the Ionosphere – (1): Ionospheric radio occultation (IRO) measurements in the limb sounding mode; (2): Use of 0.1 Hz sampled GPS navigation data from the zenith viewing antenna.

The radio occultation technique enables the retrieval of the vertical refractivity profile of a planetary atmosphere travelled by an electromagnetic wave in the limb sounding geometry. Measured is the change of ray path bending, phase or signal strength of the radio wave while approaching the planetary surface until it is completely occulted by the planet (Eshleman, 1975).

Thus, planetary atmospheres from Mars and Venus were explored by radio communication link occultations of Mariner sondes IV (*Kliore et al., 1965*) and Venera 4, respectively.

VIKING and Voyager 1 tracking and telemetry signal occultations were used to explore the ionospheres of Mars (Breus et al., 1998) and Titan (Bird et al., 1997).

In the late 1980's, when the occultation science possibilities of GPS were recognized, it was proposed to apply the radio occultation technique also to the Earth's atmosphere sounding using the L-band signals of the global positioning system GPS that was just established (*e.g. Yunck et al., 1988*). To proof this concept, the GPS/MET experiment onboard the Microlab 1 satellite mission led by the University Corporation of Atmospheric Research (UCAR) was launched in April 1994 (*e.g. Ware et al., 1996*). The GPS/MET results have demonstrated that the GPS radio occultation technique is a powerful tool for remote sensing of the Earth's neutral atmosphere and ionosphere (e.g. Kursinski et al., 1997, Hajj and Romans, 1998, Hocke and Igarashi, 2002). Consequently, several satellite missions have flown with GPS radio occultation receivers such as OERSTEDT (Escudero et al., 2001), CHAMP [Reigber et al, 2000, Jakowski et al., 2002, Garcia-Fernandez et al., 2003], and SAC-C. Since future missions will also use the signals from other Global Navigation Satellite Systems (GNSS) such as GLONASS and GALILEO, the following text refers to GNSS only.

If a GNSS receiver is installed onboard a LEO satellite, the received signals are usually used for orbit positioning and for deriving a precise board time to which the measurements are referenced. The reception of dual frequency signals from GNSS satellites at another satellite principally enables determining the total electron content between both satellites for estimating the ionosphere/plasmasphere state (e.g. Jakowski and Bettac, 1994). If the receiver is flying on a LEO satellite, the sounding covers the volume between the satellite orbit and the GNSS orbit height at about 20000 km. Using CHAMP navigation data this innovative technique has been at first applied by Heise et al. (2002). The 0.1 Hz sampled navigation data, measured with the zenith viewing antenna were effectively used for reconstructing the topside ionosphere in the vicinity of the satellite orbit plane.

For demonstrating the capabilities of both techniques as illustrated in Figure 3-21, results of the ionosphere monitoring obtained onboard the current geo-research satellite mission CHAMP (Reigber et al., 2000) are discussed. The German CHAMP (CHAllenging Minisatellite Payload) satellite was successfully launched on 15 July 2000 into a near polar orbit (inclination 87°, altitude 450 km).The satellite is equipped with a dual frequency «Black Jack» GPS receiver which enables not only the use of GPS radio occultation measurements in the limb sounding mode, but also the analysis of the 0.1 Hz sampled navigation data.

The GPS data measured onboard CHAMP are received at the DLR Remote Sensing Data Center in Neustrelitz and subsequently processed at DLR by an operational data processing system (Wehrenpfennig et al., 2001).

3.4.2   GPS Measurement Techniques

3.4.2.1   Ionospheric Radio Occultation Measurements

The radio occultation measurements rely principally on accurate measurements of the GNSS signal phases onboard a Low Earth Orbiting (LEO) satellite. The so-called phase path excess can then be used to derive the bending angle of the ray path or to determine the Total Electron Content (TEC) along the measured radio link. This measurement is the basis for retrieving the vertical refractivity profile from the LEO orbit height down to the Earth surface. Since the index of refractivity of the ionosphere depends mainly on the number of free electrons, the inversion of the measured signals can provide the vertical electron density profile (e.g. Hajj and Romans, 1998, Schreiner et al., 1999, Jakowski et al., 2002a).

The well known scheme of radio occultation is shown in Figure 3-22. The refraction angle α, between the ray path asymptotes can be derived from the GNSS carrier phase measurements onboard the LEO satellite with high accuracy. Since the bending angle is principally less than one degree, the orbit data are required with high precision (centimeter range) and clock drifts have precisely to be removed.
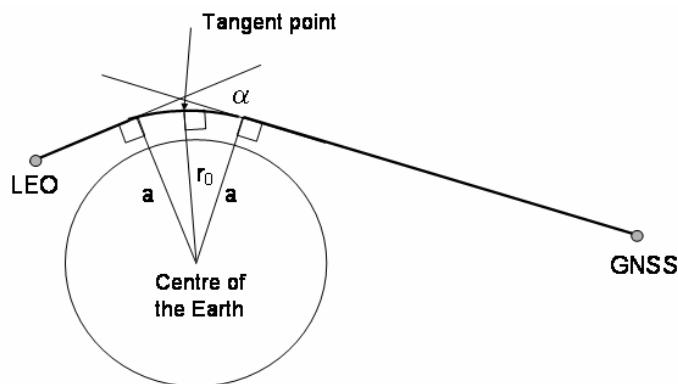
**Figure 3-22: Principal GNSS Radio Occultation Geometry for Retrieving
the Vertical Electron Density Profile of the Ionosphere.**

Introducing the impact or approaching parameter $a = n \cdot r$ that describes the refractive distance of the asymptotic ray path from the centre of the Earth, the refraction angle α can be expressed by the refraction index n via the integral equation

$$\alpha(a) = -2a \int_{r_0}^{\infty} \frac{1}{\sqrt{r^2 n^2 - a^2}} \frac{d \ln(n)}{dr} dr \;.$$

(1)

This integral equation can then be inverted by the Abel integral transform providing the vertical profile of the refractive index in terms of α and a (cf. Fjeldbo et al. 1971).

Thus, measuring the bending angle α at the refractive distance a from the satellite orbit height down to the bottom of the ionosphere, one can retrieve the vertical refractivity profile.

Taking into account L-band signal frequencies of GNSS, the refractive index can be approximated by:

$$n^2 = 1 - K \frac{n_e}{f^2}$$

(2)

with K= 80.6 $m^3 s^{-2}$, where n_e is the local electron density and f is the radio wave frequency (Davies, 1990).

Thus, inverting the integral equation (1), the vertical electron density profile below the satellite orbit height may be derived.

Oppositely to the radio occultation sounding of the neutral atmosphere, the Ionospheric Radio Occultation (IRO) measurements can take advantage of the dispersive nature of the ionosphere. Thus, differential GNSS phases derived from dual frequency GNSS measurements can effectively be used to compute the integral of the electron density along the ray path which is commonly well known as the slant Total Electron Content (TEC) of the ionosphere.

The dual frequency measurement principle is the same as applied to ground based GNSS measurements for ionospheric monitoring (e.g. Wilson et al. 1995, Jakowski, 1996).

The GPS receiver onboard CHAMP measures carrier phases in the radio occultation or limb sounding mode starting at CHAMP orbit tangential heights down to the Earth surface with a sampling rate of 1 Hz.

The dual frequency based IRO retrieval method uses only the carrier phase measurements at L_1 and L_2 GPS frequencies which are described by the observation equation:

$$\phi = \rho + c(dt - dT) - d_I + d_{MP} + dq + dQ + N\lambda + \varepsilon \qquad (3)$$

where ρ is the true geometrical range between GPS satellite and receiver, c is the vacuum speed of light, dt and dT are the satellite and receiver clock errors, d_I is the ionospheric delay along the ray path s, d_{MP} is the multipath error, dq and dQ are the instrumental satellite and receiver biases, λ is the radio wave length, N is the phase ambiguity number (integer) and ε is the residual error. The space weather sensitive ionospheric propagation term d_I is a function of the refraction index and can be written as:

$$d_I = \frac{K}{2f^2} \int_S^R n_e ds \qquad (4)$$

Here the integral of the local electron density n_e along the ray path between satellite S and receiver R is the Total Electron Content already mentioned above. Ignoring the multipath term, instrumental delays and integer ambiguities N in eq. (3), the differential carrier phases $\Delta\phi = \phi_1 - \phi_2$ computed from carrier phase measurements at L1 and L2 frequencies provide low noise TEC values as a function of the tangential height. Applying this method, dispersive ray path bending effects can be commonly ignored because these effects are small compared with the first-order-effects (Schreiner et al., 1999). The 1 Hz sampled relative TEC is measured along the radio occultation ray path which continuously approaches to the Earth surface measured by the tangential height.

The obtained measurements form a set of equations which are successively solved from top to down, to provide the electron density while assumed spherically stratified ionosphere (Jakowski et al., 2002).

Since CHAMP has a rather low, even further decreasing orbit height of less than 450 km, the upper boundary condition is ill posed. This is due to a rather high plasma density above the occultation entry. To overcome this upper boundary problem, a specific model assisted technique has been developed. The solution starts with the first measurement at the greatest tangential height by using an adaptive model for the topside ionosphere and plasmasphere above the CHAMP orbit height. The model is based on a Chapman layer function and an exponentially decaying term representing the plasmasphere. Model parameters such as the plasma scale height at the upper boundary are determined in a few iterations in order to ensure a smooth transition between model values and measurements (Jakowski et al., 2002).

To fulfill space weather monitoring requirements, i.e. to come up with retrieval products within a latency of less than 3 hours as it is required by the traditional weather service, no further data are included in the retrieval procedure and for reasons of simplicity a spherically layered ionosphere is assumed (Abel inversion assumption). The retrieval can be improved if additional information, e.g. on horizontal gradients or local densities, as may be provided by TEC maps (e.g. http://www.kn.nz.dlr.de/daily/tec-eu). On average, from about 200 IRO measurements per day, about 150 electron density profiles are successfully retrieved (Figure 3-23) now adding up to the huge number of more than 200,000 profiles.
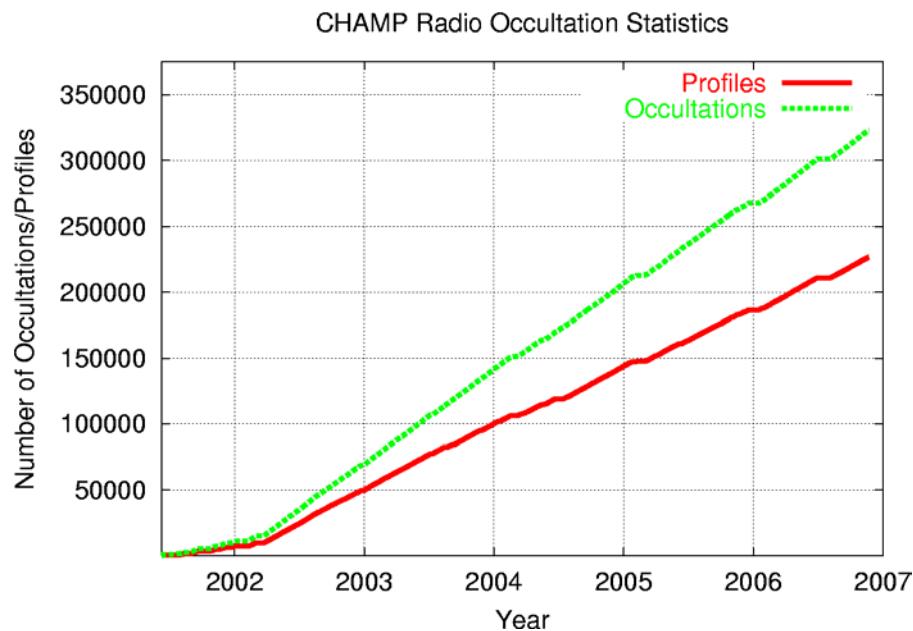
**Figure 3-23: Number of IRO Measurements and Retrieved Vertical Electron
Density Profiles from 11 April 2001 until November 2006.**

Because the processing system works automatically (Wehrenpfennig et al., 2002), the occurrence of outliers in the derived profiles cannot be avoided; however, the number of such outliers is less than 1%.

As Figure 3-24 shows, the measurements are obtained from all geographic zones, thus well suited to get global information on the actual state of the ionosphere in the context of the space weather development. Due to the nearly sun synchronized orbit of CHAMP the local time sector is slowly shifting from day to day within a repetition period of about 130 days.
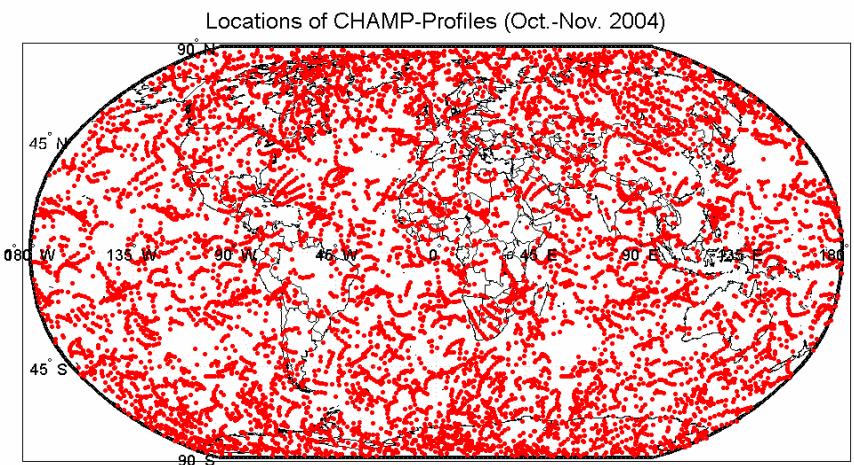
**Figure 3-24: Locations of Retrieved Radio Occultation Profiles
during Two Months (October – November 2004).**

A typical product obtained after automatic retrieval of IRO measurements onboard CHAMP at 48.93 °S; 128.76°E over the Pacific Ocean around local noon is shown in Figure 3-25. The corresponding data files can be downloaded via the DLR space weather service (http://w3swaci.dlr.de).

To estimate the quality of the derived electron density profiles, a number of validation efforts were undertaken including ionosonde data from the European vertical sounding stations Juliusruh (54.6°N; 13.4°E), Athens (38.0°N; 23.5°E), Rome (41.9°N; 12.5°E), Tortosa (40.8°N; 0.5°E) and Dourbes (50.1°N; 4.6°E) (Jakowski et al., 2005a).
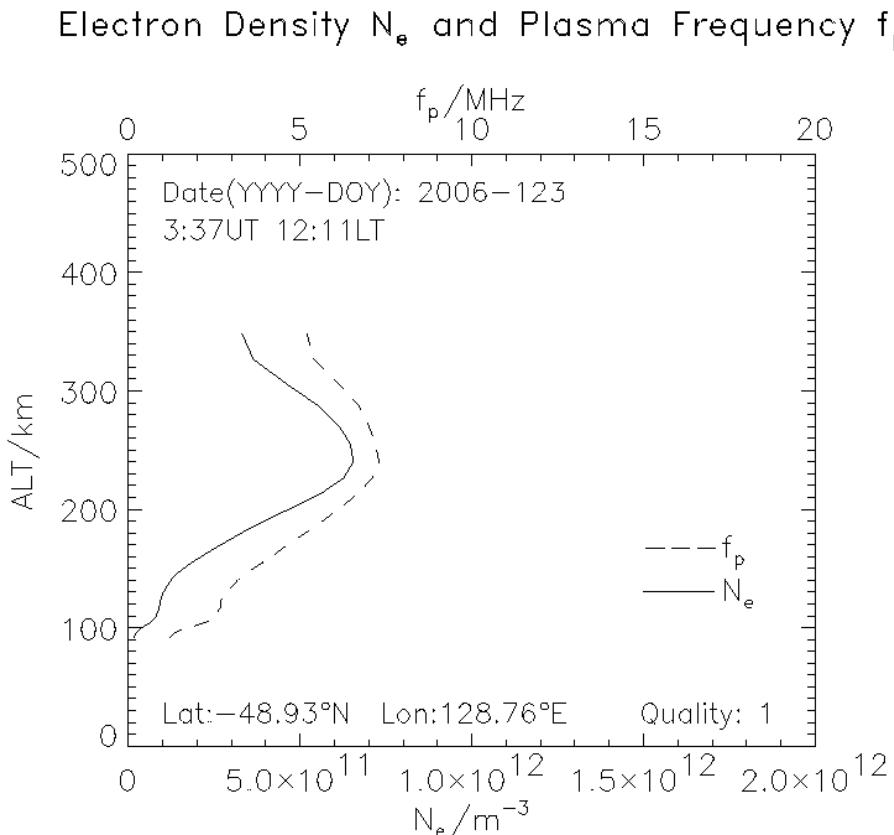
Figure 3-25: Data Product Sample Showing the IRO Retrieval of Vertical Electron Density Profiles Provided by the SWACI Service.

To give an impression of the achieved accuracy, the validation with ionograms of the Juliusruh ionosonde station has indicated a bias of up to 0.5 MHz and a RMS error of about 1 MHz in the plasma frequency (Jakowski et al., 2004, 2005a). More validation work is required in particular for low latitude data. Principally, it has to be stated that the IRO derived electron density profiles provide a unique measure describing the mean vertical electron density structure in comparably large areas with characteristic lengths of about 1000 km.

3.4.2.2   Topside Ionosphere / Plasmasphere Measurements

Whereas the IRO retrieval technique can work with non-calibrated carrier phase derived TEC data (Jakowski et al. 2002), the topside assimilation reconstruction technique requires calibrated TEC data along the numerous radio links between the GPS satellites and the topside GPS antenna onboard CHAMP (Figure 3-26). Depending on the relative CHAMP-GPS satellite constellation the data coverage changes permanently. Usually the data are not homogeneously distributed as Figure 3-26 demonstrates. To overcome this problem, the reconstruction is made via data assimilation into a reliable background electron density model.
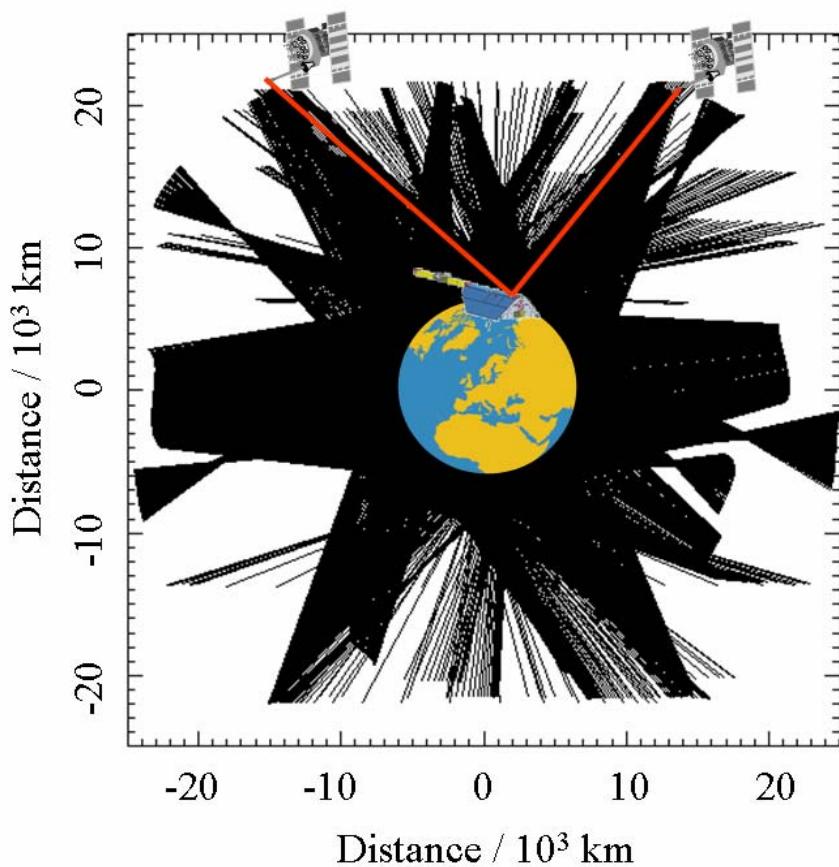
**Figure 3-26: Illustration of the Topside Radio Link Distribution in the CHAMP
Orbit Plane to the Visible GPS Satellites during One Satellite Revolution.**

Before starting the reconstruction, the corresponding satellite and receiver biases (eq. 1) have to be estimated properly before the assimilation procedure can be started.

During the pre-processing stage, detected outliers are being removed and cycle slips are corrected. For reconstructing the topside ionosphere/plasmasphere electron density the Parameterized Ionospheric Model PIM (Daniell et al., 1995) has been selected to act as the background model. After calibrating the differential phases, the absolute TEC data are assimilated into the PIM model by a method described by Heise et al. (2002). The assimilation results provide a 3D reconstruction of the electron density for each CHAMP revolution in the vicinity of the CHAMP orbit plane.

Validation of the derived electron density distribution was made with in situ plasma density measurements of the Planar Langmuir Probe installed onboard CHAMP and incoherent scatter measurements at different sites. Compared with the Langmuir Probe data the assimilation results have no significant bias and agree within a standard deviation of $2 \times 10^{11} \text{m}^{-3}$ (Heise et al., 2002). Good agreement was also found with topside profiles deduced from incoherent scatter measurements.

As Figure 3-27 demonstrates, the global view on the Earth's plasma environment enables the study of magnetospheric-ionospheric coupling processes. Here, the compression of the plasmasphere at the day-side and the enlarged extension of the plasmasphere at the night-side are clearly visible. Thus, it becomes evident that this type of space based GPS measurements can provide essential contributions to a space weather monitoring of the ionosphere.
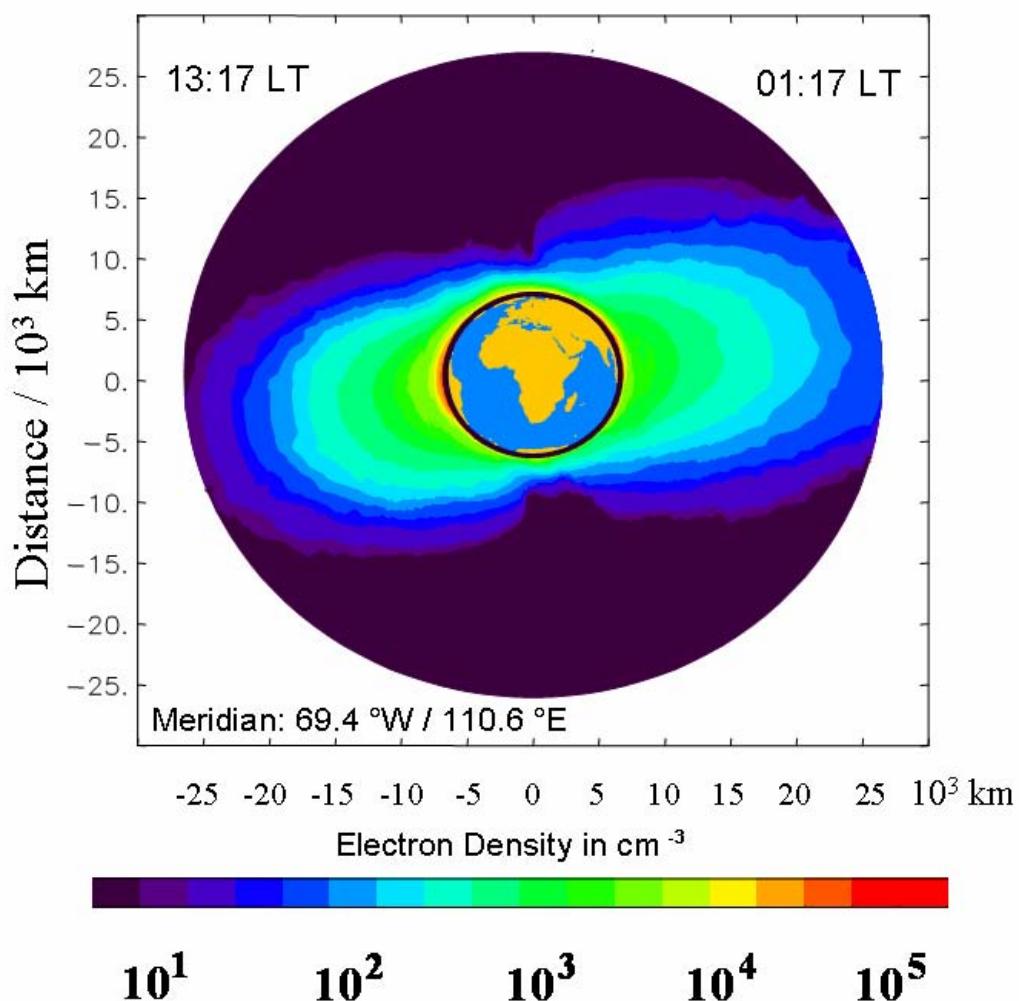
Figure 3-27: Reconstruction of the Electron Density Distribution of the Topside Ionosphere Based on GPS Data Received Onboard CHAMP.

The reconstruction is based on medians obtained for 21:00 UT over 10 consecutive days in August 2005. Thus, the right side shows the ionosphere/plasmasphere shortly after midnight whereas the left side represents ionosphere shortly after noon.

3.4.3   Observation Results and Discussion

Global data coverage and the huge amount of more than 200000 electron density profiles allow efficiently studying global large scale ionospheric processes. Because the first data were obtained in 2001 solar activity dependent effects can be studied.

As for the terrestrial weather at the Earth's troposphere, also the ionosphere is mainly controlled by the intensity of the solar radiation and its angle of incidence. Since the ionization is essentially produced by the solar ultraviolet radiation of wavelengths shorter than 130 nm, there is a strong solar cycle variation of the ionospheric peak density and other parameters related to the structure of the thermosphere/ionosphere. The solar activity control of the ionization level, well-known from TEC-measurements (e.g. Jakowski et al., 1991) is clearly visible in the peak electron density and the peak density height derived from IRO measurements onboard CHAMP (Figure 3-28).
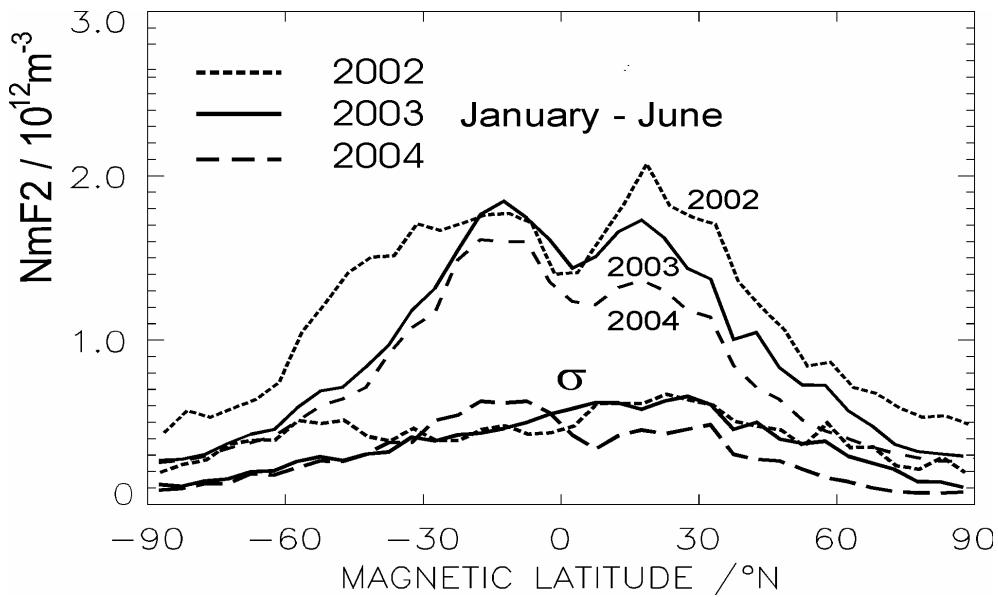
Figure 3-28: Latitudinal Dependency of the Day-Time (08:00-16:00 LT) F$_2$ Layer Peak Electron Density N$_m$F$_2$ as Seen in the CHAMP IRO Data at Three Years 2002-2004 for All Longitudes. The corresponding standard deviation σ is given at the bottom of the figure.

It is obvious that the photo ionization of the Earth's atmosphere due to solar radiation depends on the incidence angle of the irradiation. Thus, the latitudinal variation of the total ionization can easily be explained. Figure 3-28 shows that the F$_2$ layer ionization reduces up to 30 % at daytime from 2002 to 2004, when the average solar radio flux F10.7 cm decreases from about 180 down to 107 by about 40%. However, since the plasma motion is strongly influenced by magnetic und electric fields, the latitudinal dependence of the peak electron density demonstrates a more complex relationship with the solar activity. The well-known equatorial anomaly, characterized by enhanced ionization at about 15° at both sides of the geomagnetic equator is due to electric fields generated near the geomagnetic equator. The ionospheric plasma is uplifted at the geomagnetic equator via $\mathbf{E} \times \mathbf{B}$ drift and, while returning back, enhances the topside electron density at both sides of the geomagnetic equator, thus forming the so-called equatorial crest. This can nicely be seen also in the latitudinal dependence of the day-time peak density height h$_m$F$_2$ in Figure 3-29 where h$_m$F$_2$ reaches an absolute maximum of about 375 km near the geomagnetic equator. The northward shift of the maximum is due to seasonal effects because the northern summer hemisphere is stronger heated than the southern winter hemisphere.
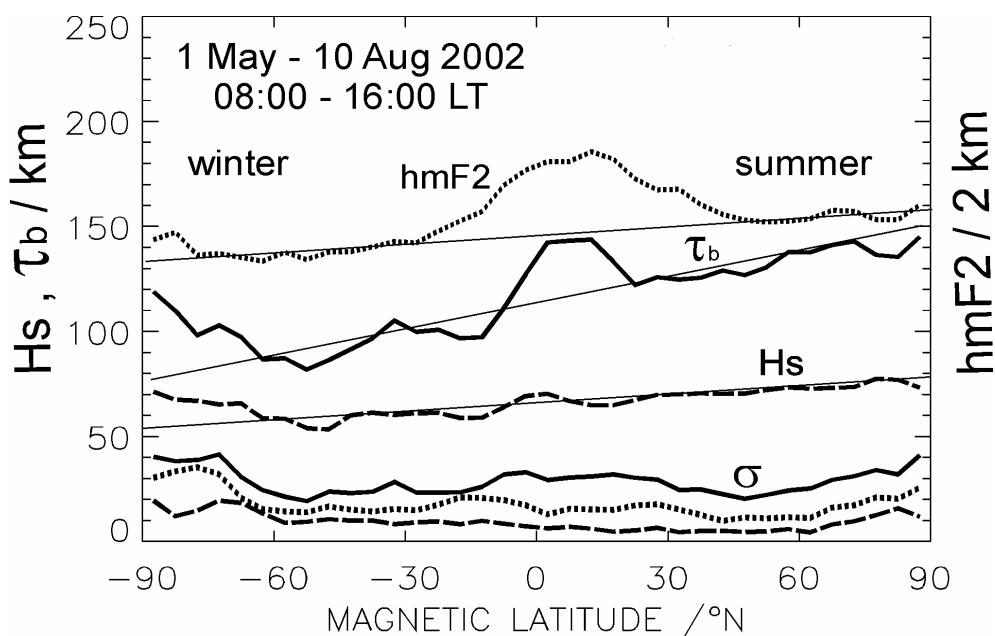
Figure 3-29: Latitudinal Dependence of the Peak Density Height h_mF_2 and Shape Parameters such as Bottomside Slab Thickness τ_b and the Topside Scale Height Hs at 425 km Height Measured at Daytime in Northern Summer. The corresponding standard deviation σ of all three parameters is given at the bottom of the figure.

Enhanced thermospheric heating during summer leads to an expansion of the thermospheric density distribution resulting in an increased peak density height and increased shape parameters slab thickness and topside scale height of the electron density profiles. This explanation is also confirmed when looking to the solar cycle dependence of these parameters showing biggest values at high solar activity conditions. If high latitudes > 65° are excluded, the general behavior indicates positive linear trends directed to the warmer summer hemisphere for h_mF_2, the scale height Hs and the bottomside equivalent slab thickness τb as well.

The following gradients may roughly be estimated: The peak density height h_mF_2 and scale height Hs grow up with about 280 m/deg and 140 m/deg, respectively, whereas the bottomside slab thickness grows up in the average by a rate of approximately 410 m/deg.

Compared with 2001, the peak density height decreases in average by about 75 km throughout all latitudes within in 2004.

The observed high latitude enhancement of all three parameters at the winter hemisphere is probably due to thermospheric heating as a consequence of particle precipitation and the action of the auroral electrojet. During night-time the polar ionosphere is separated from the mid-latitude ionosphere by the so-called trough region which is characterized by very low electron densities. As Figure 3-30 shows, IRO profiles reflect this phenomenon very well in their average plot.
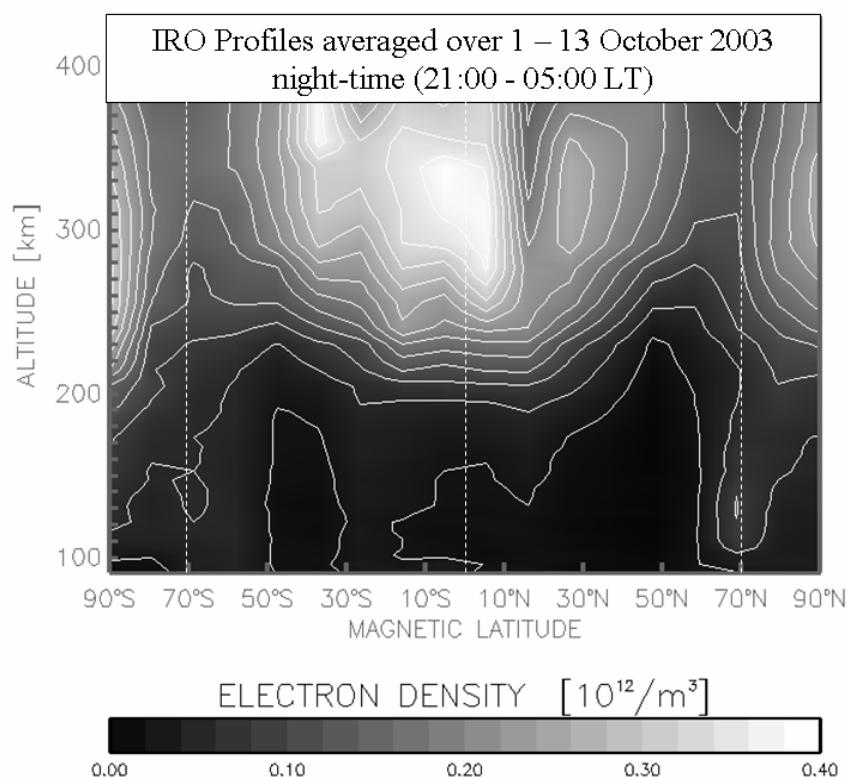
Figure 3-30: Imaging of the Average Vertical Ionization Structure Constructed from all IRO Profiles Obtained During the First 13 Days in October 2003 between 21:00 and 05:00 LT. Marked are geomagnetic latitudes at 70° and at 0°.

Although the limb sounding mode principally leads to a strong averaging of the observational data, the IRO measurements document the existence of numerous irregularities in the ionosphere (Tsybulya and Jakowski, 2005). As Figure 3-31 shows, relative irregularities of TEC with characteristic scale lengths of 15-30 km appear mainly in the nighttime sector. Here the irregularities occur just before sunrise with higher probability than near sunset. It has been shown that the disturbance pattern completely rotates with the terminator during a full day, underlining the strong relationship of the irregularity occurrence with the nighttime. Further studies are needed to explore the interaction of thermospheric conditions, e.g. atmospheric gravity waves, and related ionospheric effects well reflected in the plasma density, here measured by the relative total electron content variations along the occultation path (cf. Tsybulya and Jakowski, 2005).
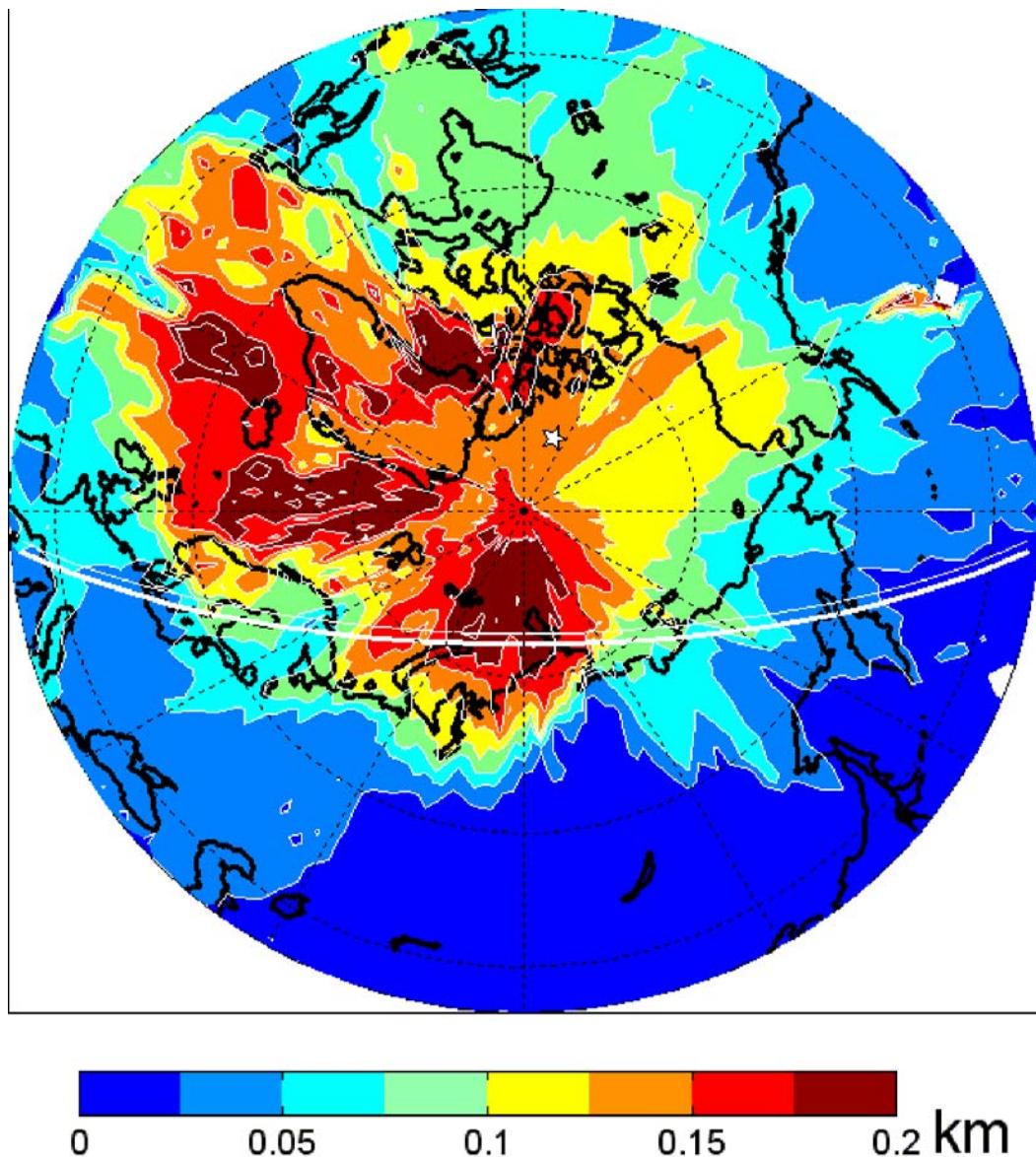
Figure 3-31: Distribution of Small Scale Irregularities of Relative TEC Measured by the IRO Technique Onboard CHAMP over the North Pole in the Winter Season (October 2003 – February 2004) Within the Time Interval 04:00-08:00 UT. The white lines mark the terminator (thick line towards the day time sector).

Severe space weather events modify the magnetosphere/ionosphere and thermosphere systems at quite different spatial and temporal scales. The zenith viewing GNSS measurements onboard CHAMP are well suited to monitor in particular large scale effects in time and space. Thus, storm induced changes of the ionospheric plasma developing at characteristic times of a few days and scale lengths of up to several thousand kilometers can be monitored effectively by analyzing the topside electron density reconstructions (e.g. Jakowski et al., 2005b).
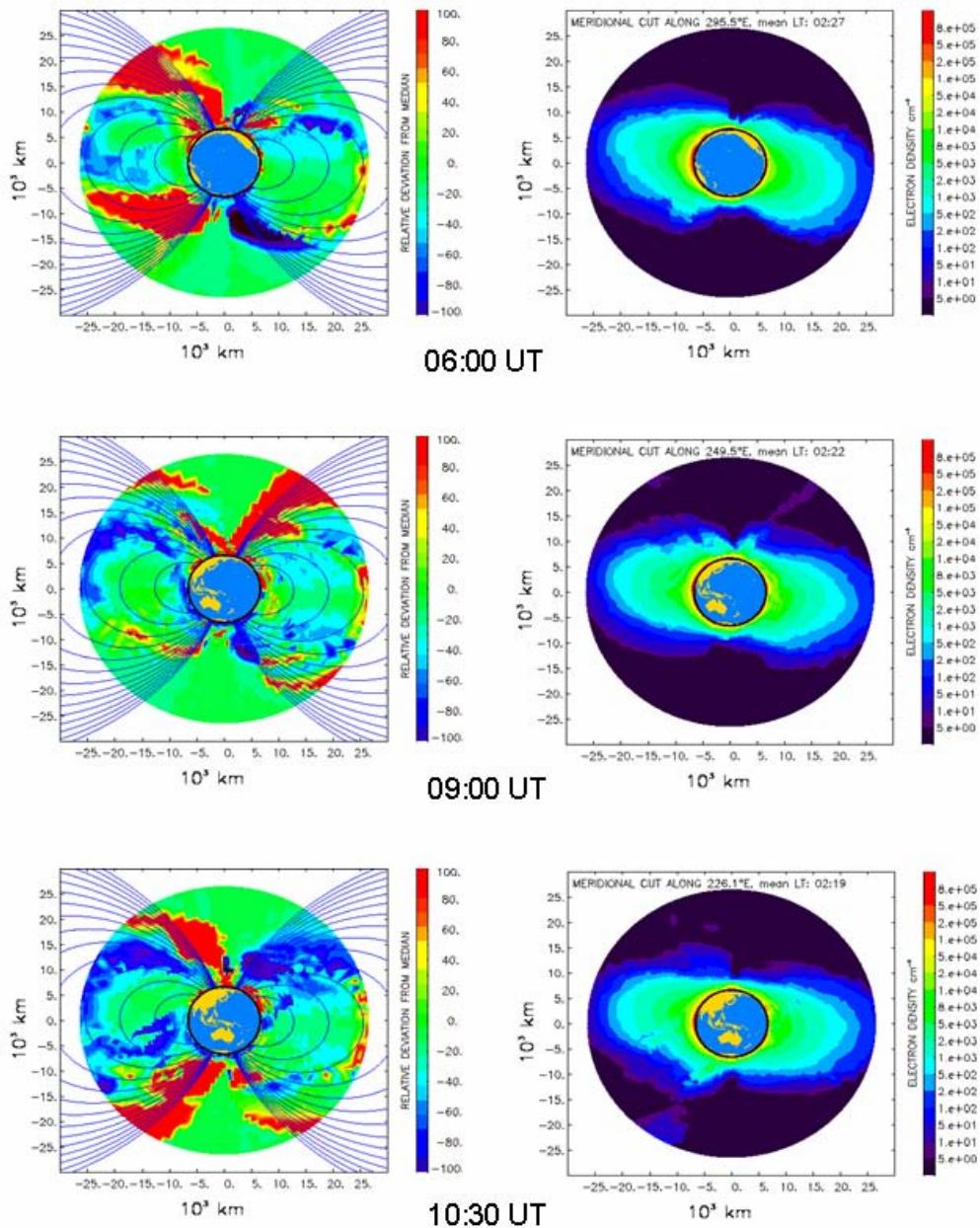
Figure 3-32: Comparison of Subsequent Reconstructions of the 3D Electron Density Structure in the CHAMP Orbit Plane during the Ionospheric Storm on 8 November 2004 in Comparison with Percentage Deviations from Corresponding Medians (Δn_e / n_e • 100%). Day-/night- sectors are on the left/right side of the figures.

During the storm on 7/8 November 2004 the electron density reconstructions in the CHAMP orbit plane indicate strong irregular behavior in particular close to the plasma pause region. The time period covered by the plots in Figure 3-32 corresponds with the time of Dst minimum value at 08:00 UT on November 8, 2004. Strong enhancements of the electron density are seen at high latitudes indicating plasma upflow in the auroral ionosphere at both hemispheres. Whereas such a plasma flow is more pronounced at the day-time polar zones at 6:00 UT and 10:30 UT, the outflow is focused at the Northern pole both at the day-

and night-time sectors around 9:00 UT. This observation agrees with the enhanced ion upflow at the polar cap boundary reported by Semeter et al. (2003) after analyzing incoherent scatter radar and IMAGE data. The cause of the upward ion velocities, which may exceed 800 m/s, remained uncertain in that study.

We are aware of the fact that the reconstructions are derived from a limited TEC data base and simplified assumptions. Nevertheless, the analysis of several storms indicates strong enhancements and also suppressions of the electron density. To avoid misinterpretation, we have counted only strong perturbation induced electron density deviations which exceed corresponding median values by 50% or are less than half of the median. The result, obtained for the storm on 20/21 November 2003, is shown in Figure 3-33.
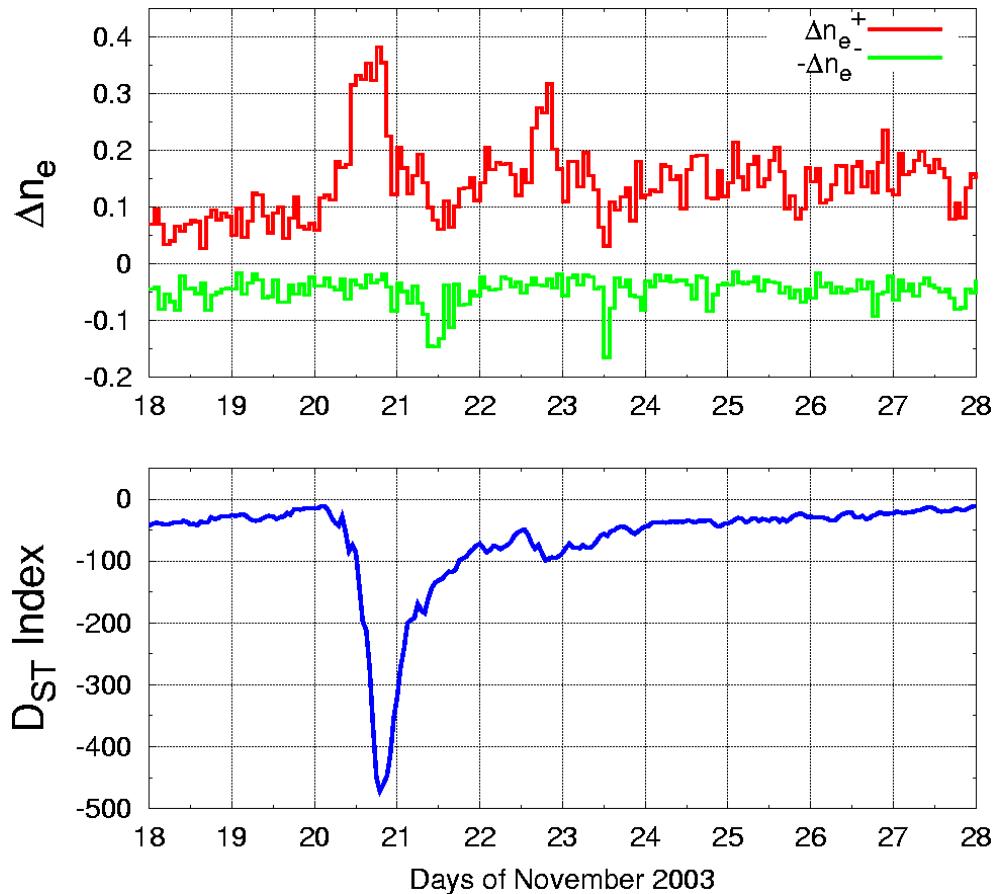
Figure 3-33: Comparison of Positive and Negative Deviation Counts of the Electron Density Values that Exceed a Certain Threshold Value ($|\Delta n_e| > 50\%$) from Corresponding Means with the Geomagnetic Dst Index (http://swdcdb.kugi.kyoto-u.ac.jp/dstdir/).

It is interesting to see that the topside ionosphere/plasmasphere considerably blows up in the growing phase of the storm whereas it reduces in the recovery phase. This behavior even repeats during a sub-storm starting around 12 UT on 22 November confirming the conclusion. Although we are aware of some remaining constraints of the reconstruction technique, e.g. due to incomplete data coverage, we believe that the basic observation results are correct.

Conclusions on the principal behavior of the plasmaspheric dynamics during geomagnetic storms are more reliable if several storms are superposed.

3.4.4   Summary and Conclusions

The large data base of IRO-derived electron density profiles and reconstructions of the topside ionosphere/plasmasphere electron density distribution is a valuable data source for the international scientific community studying ionospheric phenomena or investigation ionospheric impact on radio systems. The obtained space based GPS measurements contribute to a better understanding of solar-terrestrial relationships, and are valuable for developing and improving global ionospheric models such as IRI (e.g. Bilitza, 2003).

Due to the near-real-time processing of the satellite data the CHAMP satellite provides operational space weather information. To make this information available as fast as possible, DLR has established a permanent space weather service which provides both ground as well as space based GPS measurements and corresponding ionospheric data products.

Validation of IRO data has revealed that the F_2 layer peak electron density f_oF_2 and the corresponding height h_mF_2 agree quite well within 20% deviation. The standard deviation of vertical sounding derived electron density profiles is in the order of 1 MHz plasma frequency throughout the entire vertical IRO profile from 100-400 km height.

Although the good agreement between IRO measurement data and other types of independent measurements has proved the quality of the IRO retrieval technique, the validation process must continue because more knowledge is required in particular if the spherical symmetry assumption is violated. Further validation is also required for the topside reconstructions being capable imaging large scale structures of severe ionospheric perturbations.

To enhance the resolution of 3D reconstruction of the global ionospheric electron density distribution, space based GNSS sounding should be combined with ground based GNSS measurements.

The launch of new occultation satellites such as Formosat-3 / COSMIC (http://www.nspo.org.tw/2005e/ projects/project3/hot.shtml) and in future also SWARM and TerraSAR-X offers great promise for establishing a huge data base well suited for modeling and monitoring the electron concentration in the near-Earth-space for scientific studies and continuous space weather monitoring.

3.4.5   Acknowledgements

3.4.6   References

D. Bilitza, International Reference Ionosphere 2000: Examples of improvements and new features, Adv. Space Res. 31, #3, 757-767, 2003.

Bird, M.K., R. Dutta-Roy, S.W. Asmar, and T.A. Rebold, Detection of Titan's Ionosphere from Voyager 1 Radio Occultation Observations, ICARUS, 130, pp. 426-436, 1997.

Breus, T.K., K.Y. Pimenov, J.G. Luhmann, A. Krymsky, T. Hagfors, and W.I. Axford, Application of VIKING Radio Occultation Data to the Future Studies of the Martian Ionosphere, Adv. Space Res. , 22, No. 3, pp. 463-470, 1998.

Daniell, R.E, L.D. Brown, D.N. Anderson, M.W. Fox, P.H. Doherty, D.T. Decker, J.J. Sojka, R.W. Schunk, Parameterized ionospheric model: a global ionospheric parameterization based on first principles models. Radio Sci., 30, 1499-1510, 1995.

Escudero, A., A.C. Schlesier, A. Rius, A. Flores, F. Rubek, G.B. Larsen, S. Syndergard, and P. Hoeg, Ionospheric tomography using Oersted GPS measurements – Preliminary results, Physics and Chemistry of Earth, A26(3), pp. 173-176, 2001.

Fjeldbo, G.A., J. Kliore, and V.R. Eshleman, The neutral atmosphere of Venus as Studied with the Mariner V Radio Occultation Experiments, The Astronomical Journal, 76, 2, pp. 123-140, 1971.

Garcia-Fernandez, M., M. Hernandez-Pajares, J.M. Juan, J. Sanz, Improvement of ionospheric electron density estimation with GPSMET occultations using Abel inversion and VTEC information, J. Geophys. Res., 108(A9), 1338, doi: 10.1029/2003JA009952, 2003.

Hajj, G.A., L.J. Romans, Ionospheric electron density profiles obtained with the Global Positioning System: Results from the GPS/MET experiment, Radio Sci., 33(1), 175-190, 1998.

Heise, S., N. Jakowski, A. Wehrenpfennig, C. Reigber, H. Luehr, Sounding of the Topside Ionosphere/Plasmasphere Based on GPS Measurements from CHAMP: Initial Results, Geophysical Research Letters, 29, No. 14, 10.1029/2002GL014738, 2002.

Hocke, K. and K. Igarashi, Electron density in the F-region derived from GPS/MET radio occultation data and comparison with IRI, Earth Planets Space, 54, pp. 947-954, 2002.

Jakowski N. and H.-D. Bettac, Proposal for an ionosphere/plasmasphere monitoring system, Ann. Geophys. 12, 431-437, 1994.

Jakowski, N., B. Fichtelmann, and A. Jungstand, Solar activity control of ionospheric and thermospheric processes, J. Atmos. Terr. Phys., 53, 1125-1130, 1991.

Jakowski, N., TEC Monitoring by Using Satellite Positioning Systems, Modern Ionospheric Science, in H. Kohl, R. Rüster, K. Schlegel (eds.), *Modern Ionospheric Science*, EGS, Katlenburg-Lindau, ProduServ GmbH Verlagsservice, Berlin, pp. 371-390, 1996.

Jakowski, N., A. Wehrenpfennig, S. Heise, C. Reigber, H. Luehr, L. Grunwaldt, T. Meehan, GPS radio occultation measurements of the ionosphere from CHAMP: Early results. Geophys. Res. Lett. 29(10), doi: 10.1029/2001GL014364, 2002.

Jakowski, N., R. Leitinger, M. Angling, Radio Occultation technique for probing the ionosphere, Annals of Geophysics, 47(2/3), 1049-1066, 2004.

Jakowski, N., K. Tsybulya, J. Mielich, A. Belehaki, D. Altadill, J.C. Jodogne, B. Zolesi, Validation of GPS Ionospheric Radio Occultation results onboard CHAMP by Vertical Sounding Observations in Europe, in: C. Reigber, H. Luehr, P. Schwintzer, J. Wickert (Eds.): Earth Observation with CHAMP, Results from Three Years in Orbit, Springer, Berlin, 447-452, 2005a.

Jakowski, N., V. Wilken, S. Schlueter, S.M. Stankov, S. Heise, Ionospheric space weather effects monitored by simultaneous ground and space based GNSS signals, JASTP, 67, 1074-1084, 2005b.

Kliore, A.J., D.L. Cain, G.S. Levy et al., Occultation experiment: Results of the first direct measurement of Mars' atmosphere and ionosphere, Science, 149, pp. 1243-1248, 1965.

Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, and K.R. Hardy, Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System, Journal of Geophysical Research, 102, D19, pp. 23429-23465, 1997.

Reigber, C., H. Luehr, P. Schwintzer, CHAMP mission status and perspectives, EOS Transactions - American Geophysical Union, 81 (Suppl. to No.48), F307, 2000.

Schreiner, W.S., S.V. Sokolovskiy, C. Rocken, Analysis and validation of GPS/MET radio occultation data in the ionosphere, Radio Sci., 34, 949-966, 1999.

Stankov, S.M. and N. Jakowski, Topside ionospheric scale height analysis and modeling based on radio occultation measurements, J. Atmos. Solar-Terr. Phys. 68, 134-162, 2006.

Semeter, J., C.J. Heinselman, J.P. Thayer, R.A. Doe, H.U. Frey, Ion upflow enhanced by drifting F-region plasma structure along the nightside polar cup boundary, Geophys. Res. Lett., 30, 22, 2139, doi:10.1029/2003/2003GL017747, 2003.

Tsybulya, K. and N. Jakowski, Medium- and small-scale ionospheric irregularities detected by GPS radio occultation method, Geophys. Res. Lett., 32, L09103, doi:10.1029/2005GL022420, 2005.

Ware, R., M. Exner, D. Feng, M. Gorbunov, K. Hardy, B. Herman, Y. Kuo, T. Meehan, W. Melbourne, C. Rocken, W. Schreiner, S. Sokolovskiy, F. Solheim, X. Zou, A.R. Anthes, S. Businger, and K. Trenberth, GPS soundings of the atmosphere from low earth orbit: Preliminary results, Bull. American Meteorological Society, 77, pp. 19-40, 1996.

Wehrenpfennig, A., N. Jakowski, J. Wickert, A dynamically configurable system for operational processing of Space Weather Data. Physics and Chemistry of the Earth (C), 26(8), 601-604, 2001.

Wilson, B.D., A.J. Mannucci, C.D. Edwards, Subdaily northern-hemisphere maps using an extensive network of GPS receivers, Radio Sci., 30, 639-648, 1995.

Yunck, T.P., F. Lindal, and C.H. Liu, The role of GPS in precise Earth observation, in IEEE PLANS '88 Position Location and Navigation Symposium Record – 'Navigation into the 21st Century', IEEE Cat. No.88CH2675-7, pp. 251-258, 1988.

3.5   BROAD BAND HF MONITORING

(Stefan Hawlitschka, FGAN Germany)

3.5.1   Introduction

Historically, the military use of the HF range was expected to decrease in favour of satcoms. But with the advent of communication techniques for large symbol rate transmissions in the HF range (NATO standard STANAG 4539) military communication via HF saw a renaissance. This trend was supported by the fact that communication via HF has further advantageous features such as independence from any relay medium other than the ionosphere, potential to establish long range data or voice transmission links at rather low cost and complexity etc. Because of the growing importance of new waveforms and communication systems which continuously adapt to the quality of the transmission path present-day signal scenarios in the HF range are so complex that it is indicated for the unauthorized user to intercept frequency bands of interest with broadband equipment.

At present first HF broadband systems with bandwidths up to 5 MHz – 6 MHz are being offered by the industry worldwide. These systems however are not generally appropriate to receive signals separately which overlap in frequency and time, a signal situation which often occurs in HF. Furthermore these systems are not able to estimate the bearing values of the individual signals, but only yield a rough estimate of the true bearing of the strongest of the overlapping signals, and no estimates of the weaker

signals. And even if it would be possible to determine bearing values of the individual signals bearing errors resulting from disturbances in the transmission medium ionosphere would remain.

The aim at FGAN is to find out whether the situation just described can be significantly enhanced by resorting to advanced hardware and algorithms for separate direction finding (DF) and reception of signals overlapping in time and frequency. In order to be able to test promising approaches an experimental broadband system for this purpose, called BRAHMS (**BR**oadband **A**utomatic **H**F **M**onitoring **S**ystem) had to be realized [1,2]. The underlying concept will be sketched in section 1.2. The focus will be on the smart antenna of BRAHMS and its ability to separate signals overlapping in time and frequency via spatial filtering. This ability will be demonstrated for a live HF signal scenario in section 1.2.3. It is shown in section 1.3 that Brahms estimates the incidence angles of the arriving HF-waves very precisely but that the angles of arrival do not necessarily match the true bearings. That is due to the fact that ionospheric disturbances such as ionospheric tilts and TIDs (**T**raveling **I**onospheric **D**isturbances) for example deviate the incident waves from great circle propagation. But vice versa, because the bearing values of HF-sources estimated by BRAHMS are highly affected by ionospheric irregularities, BRAHMS is an excellent tool for observing the ionosphere. Ionospheric effects onto the direction finding with BRAHMS will be shown. By analysing the deviations of the estimated bearings from the true ones by observation of known emitters such as radio stations or weather services, TIDs and other ionospheric irregularities could be detected. As a broadband system BRAHMS receives signals of many sources at once and by the combination of the bearings of many sources with known location a spatial survey of ionospheric "distortions" shall be generated. In this way dynamic processes like TIDs shall be tracked. Finally in section 1.4 a summary as well as an outlook on the future will be given. The outlook encompasses a coming development where the spatial survey and the tracking of plasmawaves shall be used to correct for the ionospherically induced DF errors.

3.5.2   Direction Finding in the HF Range

The DF error sources in HF may be divided into five major categories: Instrumental errors are due to imperfections concerning the used hardware such as amplifiers etc. Observational errors occur at the operator interface.. Site errors are for example due to reradiators and reflectors in the vicinity of the direction finder or discontinuities in the ground characteristic etc. These errors can be minimized by proper choice of the hardware components and scrupulous choice of the receiving site. Contrary to these 'man made' DF errors, those belonging to the remaining 2 categories are due to effects occurring on the propagation path. These errors can be divided into two categories, where one category comprises those DF errors which result from multipath propagation and/or intra channel interference, while the other category includes all errors, which result from ionospheric irregularities such as tilts and TIDs.

Particularly susceptible to DF errors due to interference of HF waves are conventional HF direction finders such as NADFs (**N**arrow **A**perture **D**irection **F**inders) which are based on standard DF principles. Such NADFs are 'single-wave' systems because they interpret an incident multicomponent wave field as a single component field. They yield one bearing value which tends towards the bearing value of the strongest wave field component. This approximate bearing may have a large bias. Our approach to solve the problem of DF errors due to multipath and/or intra channel interference consists in using 'multi-wave' direction finders. Required is a synchronous multi channel data processing chain consisting of an antenna array, each element being connected to one channel of synchronous signal processing. This enables one to resolve a multicomponent wave field into the single components and to calculate a bearing value for each component.

But even if the problem of DF errors in the HF range due to multipath propagation and/or intra channel interference can be solved by applying advanced DF principles, DF errors due to ionospheric irregularities are immune against such measures. Therefore we embark on a strategy to model the ionospheric irregularities by estimating bearing values for sources with known locations such as broadcast stations and

by comparing them to the true bearings. The idea is thus to gather information for the correction of propagation induced DF errors in the case of sources with unknown locations.

3.5.3 The Experimental Broadband System BRAHMS

At present BRAHMS is being realized as an offline system, to keep the hardware effort within limits and at the same time to preserve a high degree of flexibility. This however is by no means a restriction because for the evaluation of the approach suggested in section 1.1 to improve the bearing accuracy in HF, it doesn't matter whether the tests are performed online or offline.

3.5.3.1 Concept of BRAHMS

The concept of BRAHMS is shown in Figure 3-34. Contrary to off-the-shelf conventional broadband systems for HF monitoring, BRAHMS is equipped with a 10-element array antenna, whose 10 antenna output signals are processed separately. The signal coming from an antenna element passes through an analogue bandpass filter with a bandwidth between 4 MHz and 6 MHz. Then it is A/D (**A**nalogue/**D**igital) converted and stored on a pack of hard drives. From this disk pack the 10 signals can be simultaneously fed into 10 overlap-and-add analysis FFT (**F**ast **F**ourier **T**ransform) filterbanks [1], where the frequency band to be monitored is split up into a large number of narrowband frequency channels (the number of the frequency channels corresponds to the FFT length). For each channel with signal activity super-resolution direction finding together with segmentation is performed. That means that for each signal of interest centre frequency, bandwidth, power, duration and azimuth and elevation are determined. If necessary, with this information spatial filters can be calculated to filter out the single signals. If the bandwidth of a signal of interest is larger than the bandwidth of the FFT-channels, those analysis filterbank channels, which are occupied by the signal are identified and their outputs are joined together in a synthesis IFFT (**I**nverse **FFT**) filterbank [1]. That means that by the application of analysis/synthesis filterbanks in combination with spatial filters it becomes possible, to separately extract signals, overlapping in time and frequency, from the monitored frequency band and thus to make them available for further individual signal processing, a task which will not be gone into detail here. The super-resolution bearing values themselves furthermore serve the task of DF error correction, the principal topic focused on in this paper.
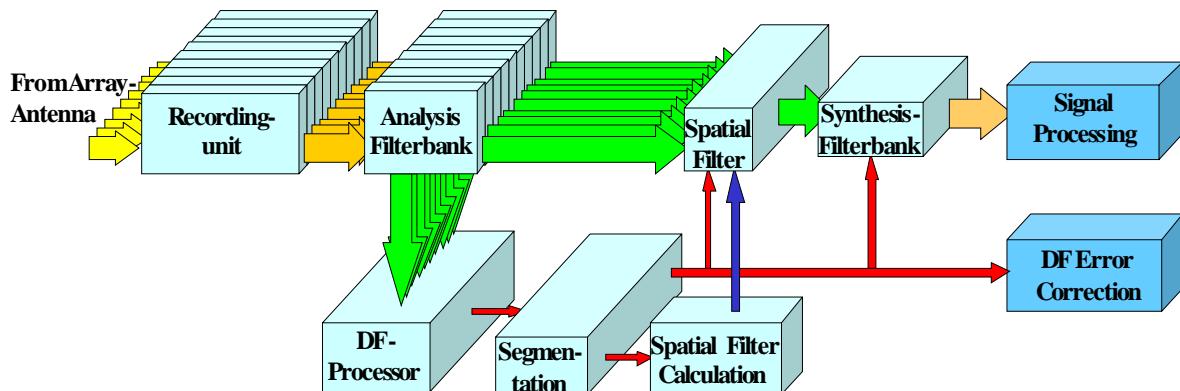
Figure 3-34: Block Diagram of BRAHMS.

At present BRAHMS can handle 2 frequency bands, one from 5,75 MHz to 10 MHz and the other from 9,085 MHz to 15,8 MHz. They have been chosen because they incorporate a number of broadcast bands and because it is our aim to use broadcast stations as calibration sources for the correction of propagation induced DF errors in the HF range. It is no problem however to add additional frequency bands by

adapting the software and by complementing the sets of analogue bandpass filters at the input of the recording unit by additional sets.

3.5.3.2   Smart Antenna of BRAHMS

The antenna elements of the 10-element array antenna are active monopoles, which are currently arranged as a double pentagon as depicted in Figure 3-35.
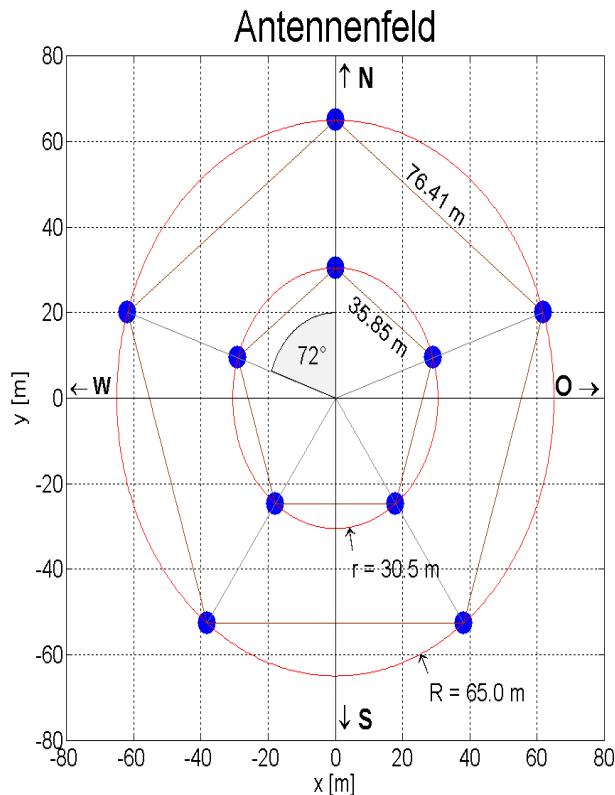
Figure 3-35: Optimized Array Geometry for Omni Directional Reception.

This geometry is optimized for omni directional reception in terms of maximum main lobe to side lobe ratio. In combination with super-resolution DF algorithms azimuth and elevation of the single components of the incoming wave field can be estimated. If a certain sector has to be reconnoitred, other array geometries such as a sparse V-shaped array geometry are preferable and if elevation is of no importance even ULAs (**U**niform **L**inear **A**rrays) may be a good choice, especially because this geometry has considerable advantages concerning the application of super-resolution DF algorithms.

Let's come back to the double pentagon array geometry. The dimensions of the array antenna shown in Figure 3-35 together with the properties of the broadband active monopoles allow to intercept HF frequency bands between 4 MHz and 15 MHz. As depicted in Figure 3-36, the beam width of the main lobe depends on the diameter of the double pentagon as well as on the frequency. As can be seen, the larger the frequency and the larger the diameter of the array, the smaller the beam width. The lower limit shown here for the beam width is the value from which on ambiguities for the array antenna arise. That means for example that a double pentagon with 10 antenna elements and a diameter of 130 m (red line) is limited to frequencies below about 15 MHz because of purely geometrical reasons.
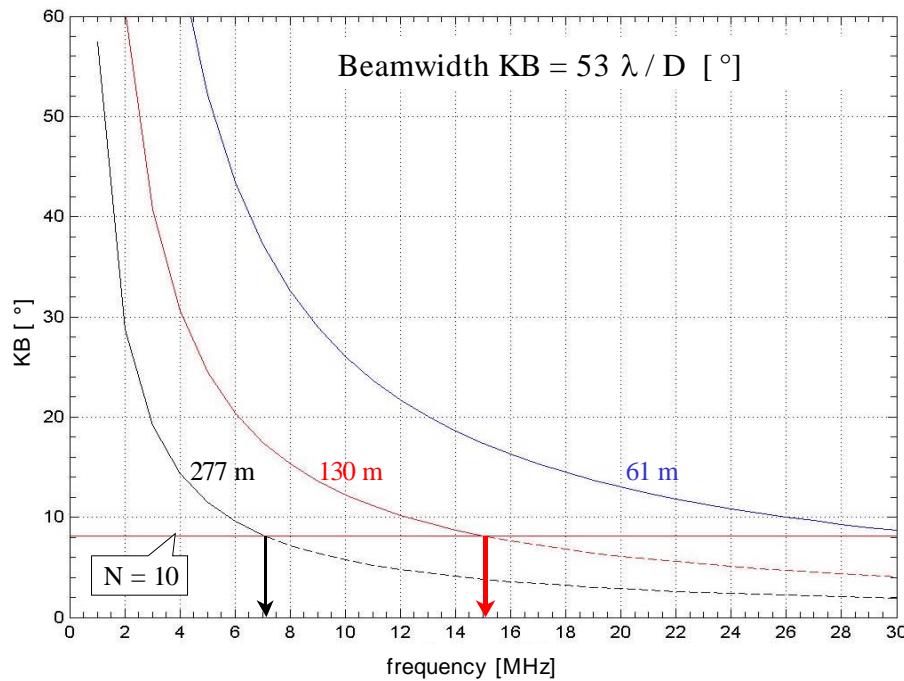
Figure 3-36: 3 dB-Beam Width versus Array Diameter and Frequency.

Next the different behaviour of conventional beamforming direction finding compared to super-resolution direction finding will be demonstrated. Figure 3-37 shows an AOA (Angle of Arrival) – diagram, where in radial direction we have the cosine of the elevation Φ, in circular direction the azimuth Θ.
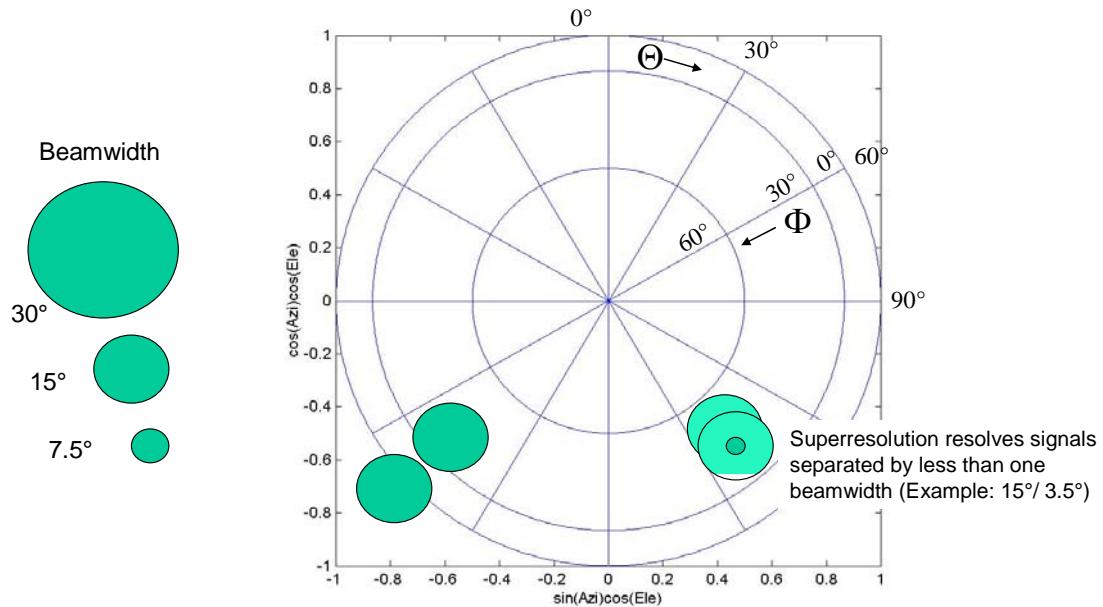
Figure 3-37: AOA (Angle-of-Arrival)-Diagram, Beam Width and Super-Resolution.

On the left side the dependence of the beam width of the main lobe on the frequency is shown: At a frequency of about 4 MHz the main lobe has a beam width of 30°, at a frequency of about 8 MHz a beam

width of 15° and at a frequency of about 16 MHz a beam width of 7.5°. Using conventional beamforming the resolution of the direction finder is about one beam width, with super-resolution and assuming a realistic HF-scenario it is about 20% of the beam width.

3.5.3.3   Algorithms

DF algorithms currently implemented on the processing unit of BRAHMS are:

- Conventional Beamforming;

- Capon or MVDR (**M**inimum **V**ariance **D**istortionless **R**esponse) [3];

- MUSIC (**M**ultiple **S**ignal **C**lassification) [4];

- Sequential MUSIC [5];

- ML (**M**aximum **L**ikelihood) [6]; and

- ML for moving sources [7].

The last one has been developed in order to take into account the ionospheric motion during the measurement interval of the AOA of an incoming HF wave. Beside these algorithms others exist which yield an analytic solution of the DF problem and for which the overall computational load is therefore much lower. Examples are:

- root-MUSIC [8]; and

- ESPRIT (**E**stimation of **S**ignal **P**arameters via **R**otational **I**nvariance **T**echniques) [9].

These algorithms have advantages over those of the first group but they presuppose certain array geometries such as ULAs for example. Therefore they are not implemented on BRAHMS at present. But BRAHMS is so flexible that it is no problem to complement its software accordingly if reasonable for example due to the use of array geometries other than the double pentagon.

Concerning the algorithms for spatial filtering, first a simple null-steering algorithm was implemented on BRAHMS. As the name suggests, this algorithm positions signals to be suppressed in nulls of the filter function. The null-steering algorithm needs bearing values to calculate the steering matrix **A** from the array manifold which is supposed to be known. Such an approach, however, requires a very good calibration of the array antennas. Furthermore the algorithm is very sensitive even to small changes of the AOAs, because the nulls of the filter are very sharp. Therefore, in a realistic HF scenario spatial filters based on the null-steering algorithm are of limited value.

Better suited is a time-frequency blind null-broadening algorithm developed at the department FE of FKIE [10]. Here also the steering matrix is used. This algorithm doesn't need either bearing estimates nor calibrated antenna arrays. It estimates the steering matrix from the output signals of the antenna array. The wave field components have to be partially non-overlapping either in frequency or time, a condition met by the HF live scenario shown on the left panel of Figure 3-38: Above estimates of the PSD (Power Spectral Density) for the time intervals [0 sec, 6 sec] and [6 sec, 8 sec] are shown in blue respectively green. Below a spectrogram of the scenario is shown. Based on the time-frequency null-broadening algorithm a spatial filter was calculated to extract the signal with the large bandwidth (green arrow) and to suppress the two others (black and blue arrows). The result is a spatial filter with broadened nulls for the robust suppression of the interfering signals which is immune against ionospherically induced slight AOA changes. Application of the null-broadening filter is depicted in Figure 3-38: The filter function is shown in the centre panel and the filtered signal on the right panel. It can be seen that the interferers are attenuated by 30-35 dB. At present work is in progress on a deterministic blind beamforming approach which even allows to separate completely overlapping signals [10].
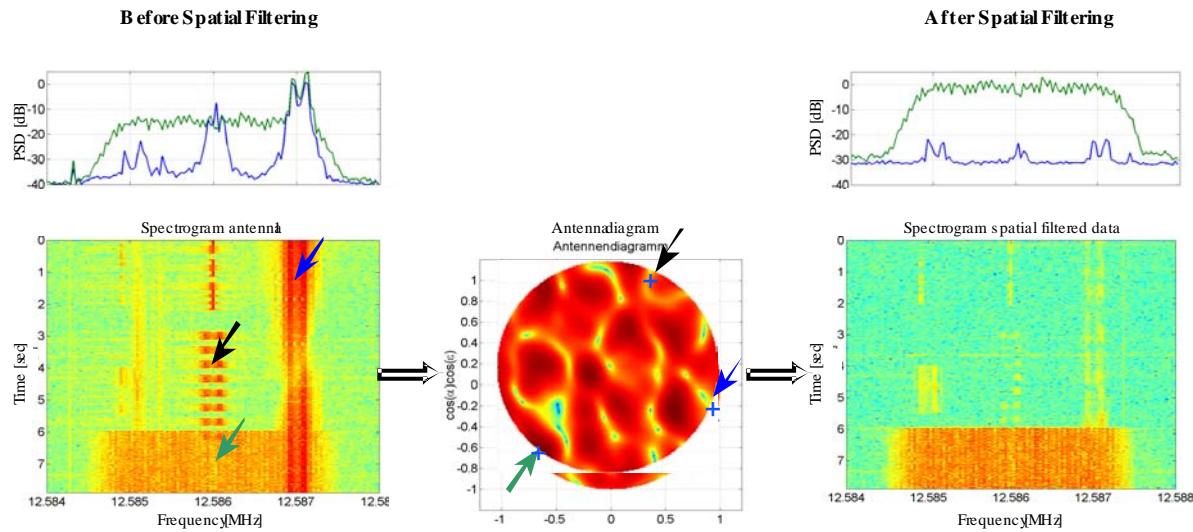
Figure 3-38: Filtering with Spatial Filter Based on Time Frequency Blind Null-Broadening.

3.5.4   Sensing the Ionosphere

So far we have presented an experimental system to obtain precise estimates of bearing angles of HF-waves onto the array antenna. By analysing the time series of the bearing angles of many emitters, information about the *actual* state of the ionosphere can be retrieved. Thus the different sources and their propagation modes (reflection on different layers) have to be separated. The approach is shown in Figure 3-39 and Figure 3-40 or an example from a live scenario. Assignment of the incidence angles of sequential measurements to a certain source and propagation mode is done by building a histogram (Figure 3-39) in the so called u,v plane (u=sin(azimuth)*cos(elevation) v=cos(azimuth)*cos(elevation)) which is segmented by applying a watershed transformation [11,12]. The result of the segmentation is shown in Figure 3-40. The different sources and their propagation modes can then be analysed individually. Limited performance may be given when the elevation angles of the two propagation modes of a single source tend to overlap. This can be minimized by reducing the sampling time down to 10 min for a histogram and it has proven to work correctly on our data. Each segment is given a number and the estimated directions are assigned to the respective segment. Then a new histogram is being sampled and segmented. The new segments, whose areas in the histogram may differ slightly are assigned to the existing ones or if a new source or mode has been detected, a new number is given. In this way data of up to three days duration have been processed.
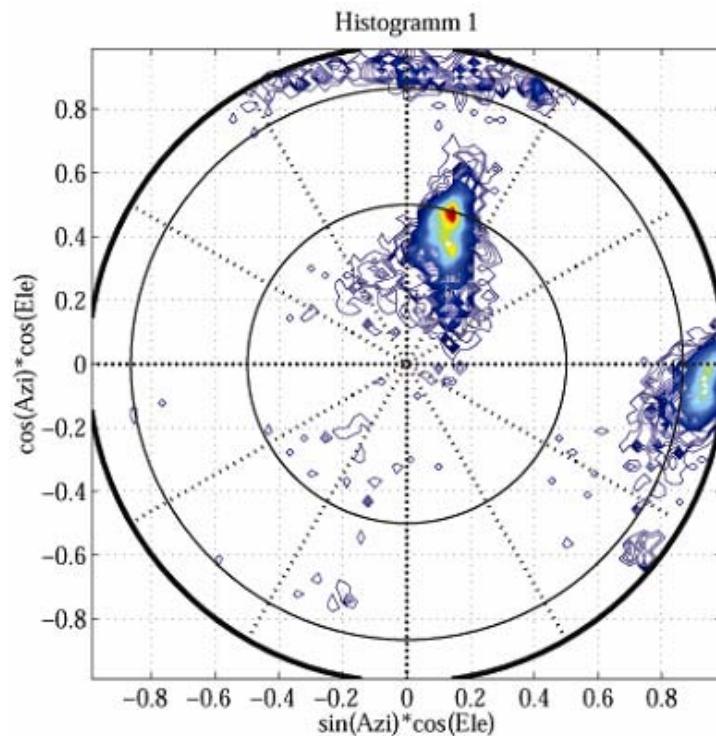
Figure 3-39: Histogram of Several Minutes of a Recording of two Main Sources, DLF Berlin with a True Bearing of 19.3° from Greding near Nuremberg and a Second Unknown Station with a True Bearing of Approximately 90°. DLF Berlin shows two peaks indicating two way propagation.
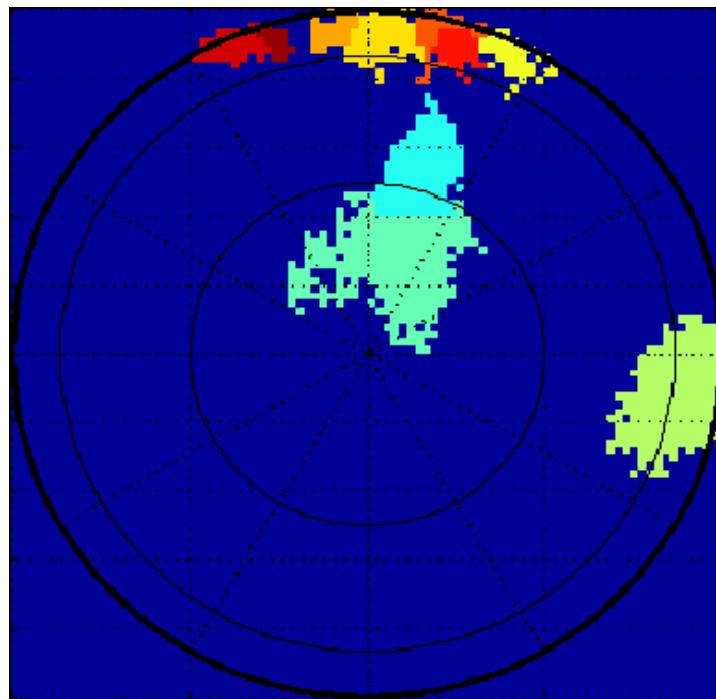
Figure 3-40: Segmentation of the Histogram of Figure 3-39. The two peaks of DLF Berlin are assigned to different segments and will be analysed separately. Also the unknown station with 90° bearing is assigned to an own segment as well as the signals coming from northerly directions.

As example a recording from the evening of 14th to the morning of 15th May 2002 is chosen which suffers from high impact of the ionosphere onto the bearing angles. In Figure 3-41 the time series of the estimated bearing angle of a transmission from the radio station DLF at Berlin is shown. It exhibits two different characteristics, being different before 23 h local summer time and after. We begin the analysis with the evening and early night part of the recording until 23 h. From 19 h to 21 h the passage of traveling wave packages (TWP) [13] can be seen. They consist of a MSTID with several periods. At 21:30 a LSTID passes through the "reflection point" of the HF wave. The corresponding first part of the signal has been segmented as one mode. The second one starts at 23 h and exhibits about 10 degrees higher elevation angles. The reason can be seen on the ionograms from Juliusruh (Figure 3-42). From 23.00 h (21:00 h UT) on spread F occurs accompanied by a second stratification 100 km above. The spread F causes higher spread of the bearing angles of the HF wave passing through it. There may still exist some LSTID action covered by the scattering in the spread F layer.
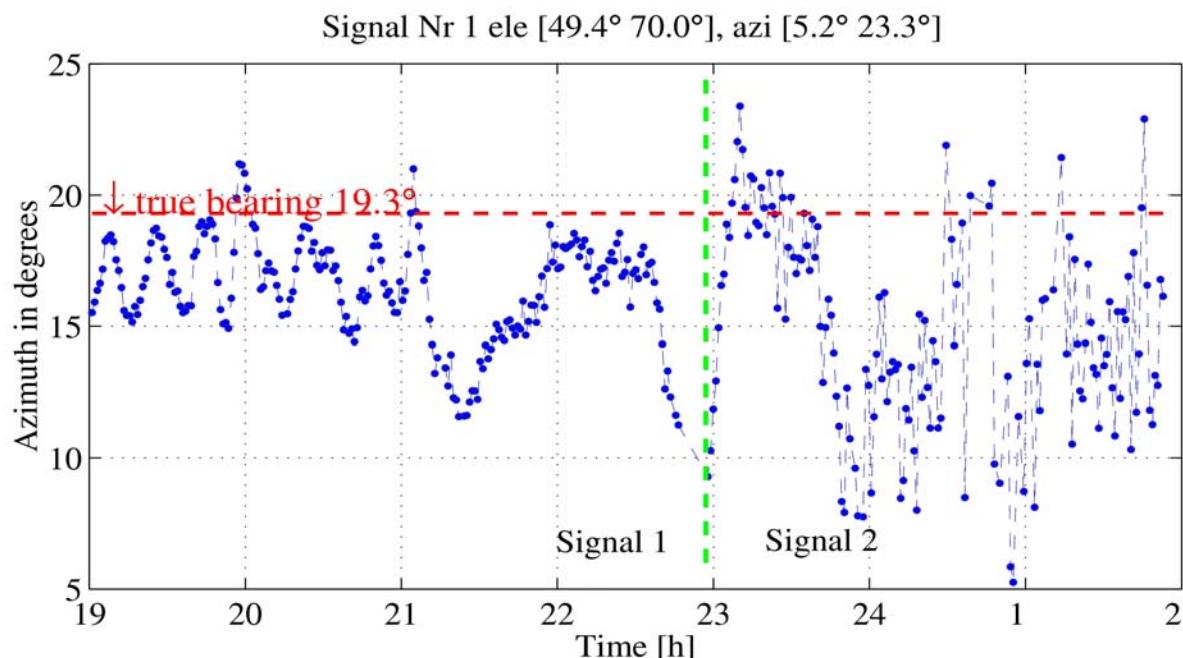
Figure 3-41: Temporal Profile of the Azimuth from the Evening into the Night. F-Region reflection of DLF Berlin on 6.005 kHz 14.05.02. The diagram is composed of Signal 1 (left of the green line) and Signal 2 (right of the green line) which has larger elevation angles. All times are local summer time (UT+2).
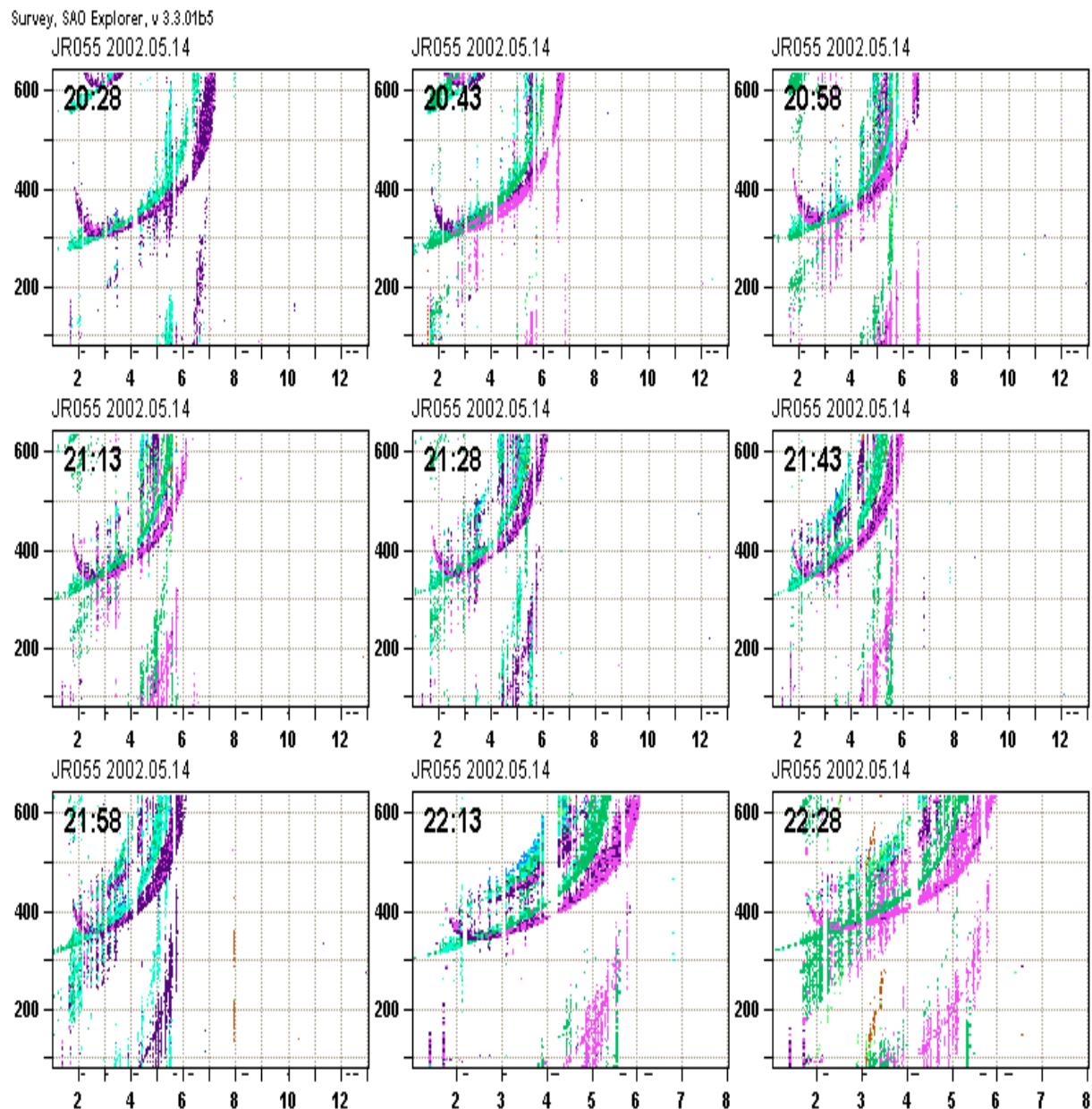
Figure 3-42: Ionograms from the Station Juliusruh / Germany for the Evening of 14.05.02. All times are UT. Shown are traces of the ordinary wave (green) and extraordinary wave (magenta).

On the next morning no more spread in the F-region propagation is detected, but TID action can be recognized (Figure 3-43). Additionally there exists a second propagation mode enabled by reflection on the E-layer (Figure 3-44). Again these two modes have been discriminated because of different elevation angles. In Figure 3-45 the mean absolute bias (blue) and the standard deviation (red) around the (biased) mean value of the data of Figure 3-44 are shown. Thanks to appropriate signal processing the spread around the mean value is quite small. The standard deviation is as in most cases smaller than the bias.
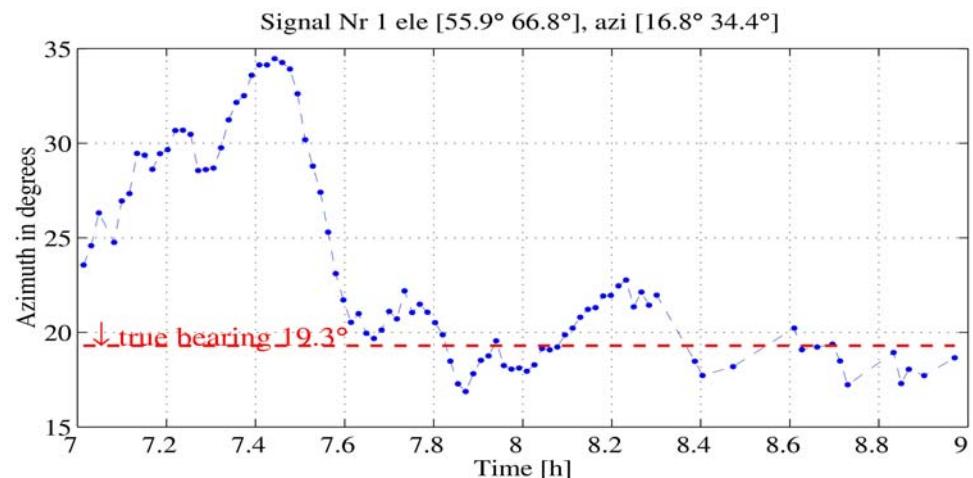
Figure 3-43: F-Region Bearing Angles on the Morning of 15.05.02 for DLF Berlin.
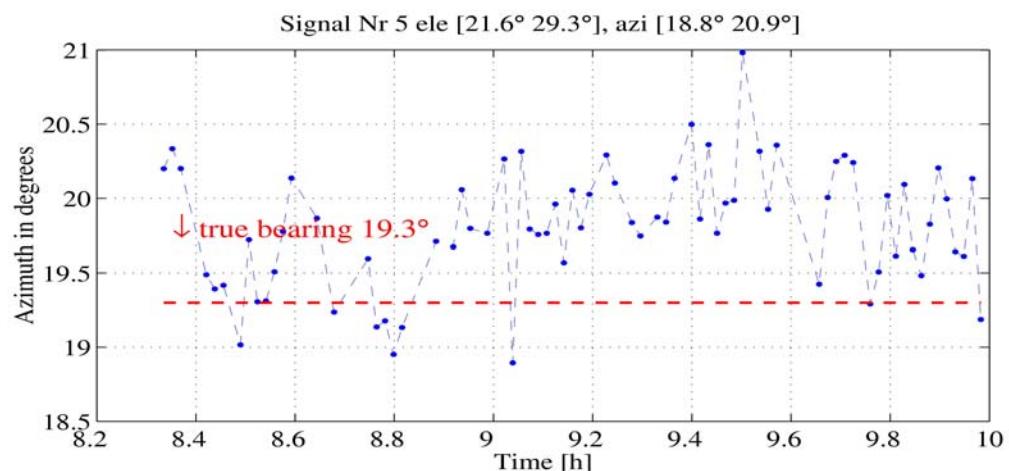
Figure 3-44: E-Region Bearing Angles on the Morning of 15.05.02 for DLF Berlin.
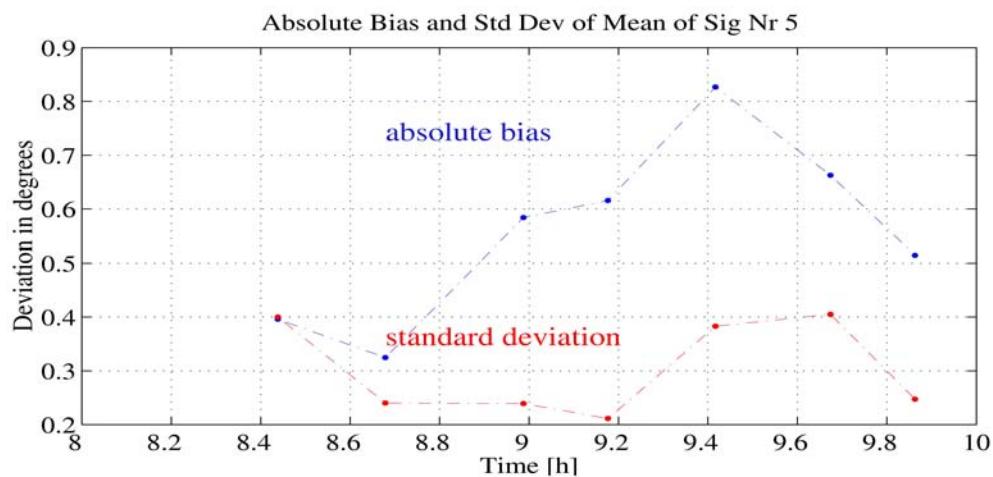
Figure 3-45: Mean Absolute Deviation (Absolute Bias) Each with 10 Samples and Standard Deviation Around the (Biased) Mean Value of the Data of Figure 3-44.

3.5.5   Summary and Outlook

The experimental system BRAHMS and its technical properties as well as its advantages for surveillance of HF have been described. It is a system for simultaneous and continuous surveillance of 4 or 6 MHz bandwidth. Unique are essentially two topics: Separate direction finding and reception of signals overlapping in time and frequency. This has been successfully demonstrated for HF signals from live scenarios.

Another highlight shall consist in a correction of DF estimation errors. As shown the short term errors have already considerably been reduced through the development of MLMS direction finding algorithms. The reduced spread of bearing angles is an indispensable prerequisite for a successful correction of the remaining DF errors. The long term DF deviations are mainly ionospherically induced. To cope with them a correction method is under development which shall be based on the actual measurement of the ionospheric state by the analysis of ionospheric irregularities as shown in Section 1.3. Many scientific analyses have been performed to characterise their effect onto DF [14,15]. Beley et al. [16] have developed a geometrical and a statistical approach to reconstruct the parameters of a TID. A similar formalism could be chosen to "translate" the deviations of the estimated bearings from the true bearings of emitters with known location, mainly from the radio bands into the shape of the ionospheric irregularities. By observation of many radio stations an image of the "undulations" of the ionosphere shall be constructed. The aim is to use this image and the knowledge about its development to correct the bearing angles of unknown sources, which have to be reconnoitred.

3.5.6   References

[1]   Hißen, H. & Baum, D. (2004). *Automatische Breitbanderfassung mit intelligenten Antennen im Kurzwellenbereich*. FKIE-Report No. 75.

[2]   Baum, D. & Foerster, W. (2006). *BRAHMS – Ein breitbandiges automatisches HF-Monitoring-System*. FKIE-Report to be published.

[3]   Capon, J. (1969). *High-Resolution Frequency-Wavenumber Spectrum Analysis*. Proc. IEEE., vol. 57, pp. 1408-1418.

[4]   Schmidt, R.O. (1986). *Multiple Emitter Location and Signal Parameter Estimation*. IEEE Trans. AP, vol. 34, pp. 276-280.

[5]   Oh, S.K. & Un, C.K. (1993). *A Sequential Estimation Approach for Performance Improvement of Eigenstructure based Methods in Array Processing*. IEEE Trans. SP, vol. 41, pp. 457-463.

[6]   Ksienski, A.A. & McGhee, R.B. (1968). *A Decision Theoretic Approach to the Angular Resolution and Parameter Estimation Problem for Multiple Targets*. IEEE Trans. AES, vol. 4, pp. 443-455.

[7]   Baum, D., Foerster, W. Hawlitschka. S. (2007): *BRAHMS Broadband Automatic HF-Monitoring System*, RTO IST-056 / RSM002 Symposium on Characterising the Ionosphere, 12-16 June 2006, Fairbanks USA.

[8]   Barabell, A.J. (1983). *Improving the Resolution Performance of Eigenstructure based DF Algorithms.* Proc. ICASSP, pp. 336-339.

[9]   Paulraj, A., Roy, R. & Kailath, T. (1986). *A Subspace Rotation Approach to Signal Parameter Estimation*. Proc. IEEE, vol. 74, pp. 1044-1045.

[10] Demissie, B. (2006). *Verfahren zur Raumfilterung für das Breitbanderfassungssystem BRAHMS*. FKIE-Report to be published.

[11] Vincent, L. & Soille, P. (1991): *Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations.* IEEE Trans. PAMI, vol. 13, pp. 583-598.

[12] Najman, L. & Schmitt, M. (1996): *Geodesic Saliency of Watershed Contours and Hierarchical Segmentation.* IEEE Trans. PAMI, vol. 18, pp. 1163-1173.

[13] Afraimovich, E.L., N.P. Perevalova and S.V. Voyeikov (2003): *Traveling wave packages of total electron content disturbances as deduced from global GPS network data.* JASTP 65, 1245-1262.

[14] Hawlitschka, S. (2006): *TIDs and Tides observed by a super-resolution Direction Finding System.* JASTP 68, 568-577.

[15] Hawlitschka, S. (2007): *Plasmawaves observed by a super-resolution DF system.* RTO IST-056 / RSM002 Symposium on Characterising the Ionosphere, 12-16 June 2006, Fairbanks USA.

[16] Beley V. S., Galushko, V. G., Yampolski, Y., M. (1995): *Traveling ionospheric disturbance diagnostics using HF signal trajectory parameter variations.* Radio Science, Vol. 30, 1739-1752.

Chapter 4 – SPONTANEOUS RADIO FREQUENCY EMISSIONS FROM NATURAL AURORA

J. LaBelle

Dartmouth College, Department of Physics and Astronomy, Hanover, NH, USA

June 9, 2006

Abstract. At high latitudes, suitably sensitive radio experiments tuned below 5 MHz detect up to three types of spontaneous radio emissions from the Earth's ionosphere. In recent years, ground-based and rocket-borne experiments have provided strong evidence for theoretical explanations of the generation mechanism of some of these emissions, but others remain unexplained. Achieving a thorough understanding of these ionospheric emissions, accessible to ground-based experiments, will not only bring a deeper understanding of Earth's radio environment and the interactions between waves and particles in the ionosphere but also shed light on similar spontaneous emissions occurring elsewhere in Earth's environment as well as other planetary and stellar atmospheres.

1. Introduction

While it cannot be excluded that radio frequency emissions might occur in all ionospheric regions, only from the auroral ionosphere, at magnetic latitudes from 65 to 75 degrees, are such emissions regularly observed and extensively documented. The auroral region uniquely contains the energy source for such emissions, which is the highly nonthermal electron distribution functions associated with the northern and southern lights phenomena.

As discussed in the Introduction, the auroral ionosphere is dominated by processes imposed upon it from the magnetosphere. In particular, magnetospheric processes drive field-aligned currents into and out of the ionosphere that must be closed across the ionospheric plasma. In the region where an upward current is imposed, this current must be carried by downward electrons. Because these are scarce at high altitudes, a field-aligned electric field develops, typically at altitudes of several thousand kilometers, to accelerate these electrons to energies of 1–10 keV, enabling them to carry the required current. As a result, at ionospheric altitudes the distribution of electrons not only includes the cold (0.1 eV), approximately isotropic ionospheric plasma, but also includes the 1–10 keV beam, which by the time it reaches the ionosphere has been scattered into a ring distribution. Electrons with velocities directed too nearly parallel to the magnetic field are lost in the atmosphere and not backscattered, so the energetic distribution actually resembles a ring with a loss cone removed from it in the upward direction parallel to the magnetic field; this distribution is sometimes called a "horseshoe." In addition to the cold plasma and the energetic beam, there is a background of warm secondary scattered particles generated by impact ionization. Detailed features result because some electrons are trapped between the converging magnetic field below and the downward-accelerating electric field above. This auroral electron distribution function was first described in a classic paper by Chiu and Schulz (1978). Taken together, these populations of electrons comprise a highly non-Maxwellian distribution, containing ample free energy. In its attempt to relax to a thermal distribution, the plasma excites both electrostatic and electromagnetic waves; the latter may propagate long distances as radio emissions, and the former may mode-convert to electromagnetic waves observable at long distances.

Other regions of the aurora may also provide sources of spontaneous emissions. In the region where a downward field-aligned current is imposed on the ionosphere, electrons are extracted and the electron density becomes depleted in the ionosphere; eventually a dearth of ionospheric electrons leads to formation of electric fields parallel to the magnetic field accelerating electrons upward in this region. In the dayside auroral latitudes near noon local time, the "cusp" feature of the geomagnetic field allows direct access of solar wind particles which provide another population of electrons. In other

auroral regions, time dependent field-aligned currents imposed by magnetospheric activity or magnetosphere-ionosphere interactions are carried by Alfvén waves which accelerate electrons locally, causing bursts of electrons directed along the magnetic field line with a broad range energies from a fraction of an eV to hundreds of eV. The electron distributions in these regions is different from that described above but is also highly non-Maxwellian, containing free energy for spontaneous emissions.

At mid- and low-latitudes, the ionosphere contains non-Maxwellian electron distributions as well, resulting for example from photo-ionization. However, the non-Maxwellian features of the auroral ionosphere, described above, are much more extreme. For this reason, the auroral ionosphere generates multiple types of spontaneous emissions which are easy to observe, whereas spontaneous emissions have not been reported from the other regions. Section 2 review the spontaneous emissions from the auroral ionosphere which are observable at ground level.

2. Emissions Observable at Ground-Level

Figure 1 shows 0-5 MHz signals recorded at Arviat, Nunavut, on April 24, 1995, using a magnetic loop antenna and a 10-kHz bandwidth stepped-frequency receiver. (See Weatherwax et al. (1995) for a description of the receiver and sensor.) This spectrogram covers 0-5 MHz on the vertical axis, 0620–0720 UT on the horizontal axis, and the gray scale represents the power spectral density ranging from 5 nV/m$\sqrt{\text{Hz}}$ (white level) to 500 nV/m$\sqrt{\text{Hz}}$ (black level). The spectrogram covers an hour around local magnetic midnight (0635 UT) which is the most favorable time for observing spontaneous emissions from the aurora. On such a spectrogram, man-made signals, which are typically constant-frequency, stand out as horizontal dark bands. The abundance of such signals at 550–1600 kHz corresponds to the AM broadcast band (since extended to 1700 kHz). However, three types of signals stand out as distinct from the man-made signals. These represent the three types of radio emissions of auroral origin: Auroral Hiss, a broad-band impulsive emission extending from VLF up to about 1 MHz; Auroral MF-burst, a broad-band impulsive emission often coincident with hiss which occurs in the range 1500–4500 kHz; and Auroral Roar, a narrow-band emission which occurs near 3 MHz and 4.5 MHz. (The 4.5 MHz emission is not shown in Figure 1.)

2.1. AURORAL ROAR EMISSIONS

Figure 1 shows an example of an auroral roar emission occuring at 2.8–3.0 MHz, which is more or less continually present from 0625–0645 UT but becomes intermittent after that. A substorm onset occurs at 0652 UT,

Apr 24, 1995   Arviat (71.4˚ inv)
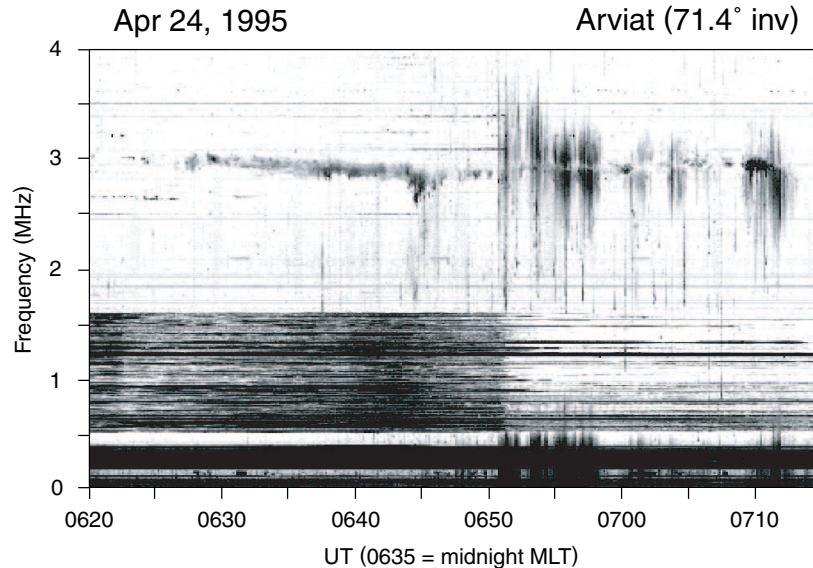
Figure 1. Auroral emissions recorded at Churchill, Manitoba (69.2° magnetic latitude) around the time of a substorm onset at 0652 UT on April, 24, 1995, showing the three types of spontaneous radio emissions generated in the auroral ionosphere and observable at ground-level: The narrowband emission centered at 2.9-3.0 MHz is auroral roar, the broadband emission at 1.5-3.8 MHz is auroral MF-burst, and the broadband emission extending up to 0.5-1.0 MHz is auroral hiss.

evidenced by the sudden decrease in the intensity of the AM broadcast signals reaching the receiver from distant sources. These broadcast signals are attenuated after the substorm onset due to absorption caused by the sudden and widespread ionization in the D-region associated with the enhanced substorm auroral activity. Typically auroral roar occurs continuously preceding the auroral substorm, sometimes for an hour or more rather than 20 minutes as in this example. After the substorm, the emission most likely continues but is probably screened from ground level by the enhanced D- and E-layer ionization. This hypothesis explains the intermittent nature of the emissions following substorm onset, but no suitably sensitive rocket- or satellite-borne experiment has yet confirmed it.

This auroral roar emission near 3 MHz is sometimes called "$2f_{ce}$-roar" because it matches twice the electron cycltron frequency ($2f_{ce}$) at F-region altitudes. This emission was first reported by Kellogg and Monson (1979). As mentioned above, there exists a second type of auroral roar which occurs

at 4.0–4.5 MHz, sometimes called "$3f_{ce}$-roar," first reported by Weatherwax et al. (1993). Observations from many observatories covering a range of magnetic latitudes show that the emission frequencies increase linearly with the magnitude of the geomagnetic field at the observatories, and the best linear fit implies that the emission frequency matches $2f_{ce}$ at about 275 km for the $2f_{ce}$-roar and matches $3f_{ce}$ at an equal or somewhat higher altitude for the $3f_{ce}$-roar (Hughes and LaBelle, 1998). At ground level the power spectral density of these emissions ranges up to 10^{-12} V^2/m^{2}Hz, which implies amplitudes of 10–100 μV/m. An isotropic radiator of 1–10 W at F-region altitudes suffices to explain such a weak signal level assuming no significant absorption. Hence, if the source size is reasonably large, greater than hundreds of meters as discussed below, the emission mechanism need not be very efficient, since 1-10 W is small compared to the auroral energy input of order 100 mW/m^2 over a source region of reasonable size. The polarization of the auroral roar has been measured and corresponds to O-mode propagation in the ionosphere (Shepherd et al., 1997).

Figure 2 illustrates the mechanism that has been put forward to explain auroral roar emissions. Altitude is shown on the vertical axis and frequency on the horizontal axis. The curves show "typical" profiles of the electron plasma frequency (f_{pe}) and the upper hybrid frequency (f_{uh}), which peak in the F-region at a few hundred kilometers altitude, and the electron cyclotron frequency and its first two harmonics which are decreasing functions of altitude. Depending on how large the electron density happens to be, locations may exist where f_{uh} matches $2f_{ce}$ or, under high-density conditions, $3f_{ce}$. These special locations are indicated by dots in Figure 2. Ample theoretical work (e.g., Kaufmann, 1980; Yoon et al., 1996, 1998a) shows that in the presence of the auroral electron distribution described above, electrostatic upper hybrid waves are favorably generated when the upper hybrid frequency matches a cyclotron harmonic. Hence, at the special locations indicated in Figure 2, one expects intense electrostatic waves in the presence of auroral electrons. If these electrostatic waves convert to electromagnetic waves, and if no blanketing E-region intervenes, they can penetrate to ground level to be observed as auroral roar emissions near 2.5–3.0 MHz and 4.0–4.5 MHz. (Gough and Urban (1983) first suggested this mechanism to explain the $2f_{ce}$ emission.) Direction finding observations of auroral roar at Sondrestrom, Greenland, show that the auroral roar signals, ray-traced into the ionospheric electron density profile observed with the incoherent scatter radar, reach points in the ionosphere where $f = 2f_{ce} = f_{uh}$ as predicted by this generation mechanism (Hughes and LaBelle, 2001b).

This mechanism predicts the existence of matching conditions at two special locations on the topside ionosphere, which should radiate electromagnetic waves upward into the topside ionosphere. Indeed, such topside emissions have been observed. From satellite data, James et al. (1974) reported
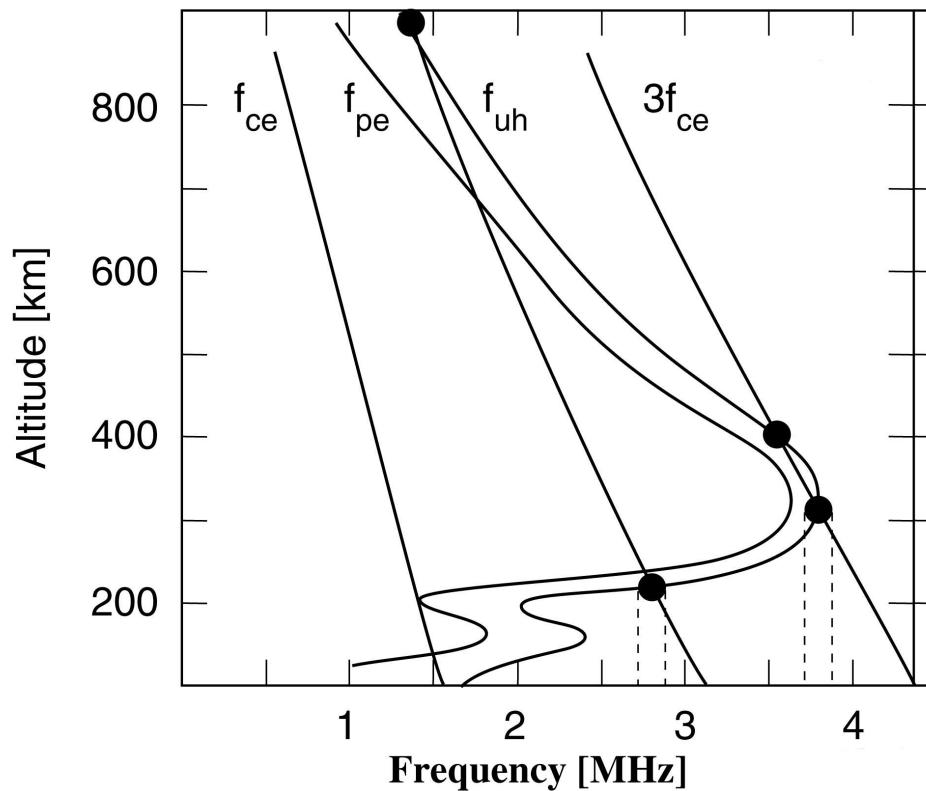
Figure 2. Typical profiles of critical resonance frequencies in the auroral ionosphere relevant to the generation mechanism of auroral roar emissions: The electron plasma and upper hybrid frequencies peak in the F-region, while the electron cyclotron frequency and its harmonics decrease with altitude. For sufficiently high F-region electron densities, locations exist where the upper hybrid frequency matches a cyclotron harmonic, indicated by dots in the figure. Intense electrostatic upper hybrid waves can be excited by the auroral electrons under these conditions. It they mode-convert to electromagnetic LO-modes, these can propagate to ground level and be observed as auroral roar.

emissions near 2- and 4-MHz in the topside ionosphere. These frequencies are somewhat below the corresponding frequencies observed at ground level, as expected from the model (Figure 2). Furthermore, from ray-tracing the observed signals into the topside ionospheric density profile determined by an ionospheric sounder on board the same spacecraft, James et al. (1974) concluded that the signals originated where f_{uh} matches $2f_{ce}$ and $3f_{ce}$. Benson and Wong (1987) analyzed many more of these topside 2- and 4-MHz signals observed with the ISIS-1 low-Earth orbit satellite, and Bale (1999)

reports one possible example observed at much greater distance from Earth with the WIND satellite.

However, the strongest evidence for the generation mechanism of auroral roar comes from sounding rocket observations. Long ago the Porcupine rocket experiment penetrated the F-region ionosphere and showed the prevalance of intense electrostatic waves occuring where $f_{uh} = 2f_{ce}$, both in the natural auroral ionosphere and in an plasma artificially enhanced by a Barium cloud explosion within the aurora [Carlson et al., unpublished manuscript, 1987]. More recently, the HIgh Bandwidth Auroral Rocket (HIBAR) encountered electrostatic waves occurring where $f_{uh} = 2f_{ce}$ near 300 km altitude in the natural auroral ionosphere (Samara et al., 2004). Unlike the earlier experiment, HIBAR included high-resolution measurements of the electron distribution function up to tens of keV with 30 ms cadence. Using these experimentally determined distribution functions, it was possible to calculate the growth rate of the Z-mode cyclotron maser instability at each frequency every 30 ms throughout the flight. The results show that strong growth rates, defined as e-folding times less than about 0.1 ms, occur only in the regions near the matching condition $f = f_{uh} = 2f_{ce}$ in which the enhanced electrostatic waves were observed; in other regions away from the matching condition, or in regions where the matching condition was met but no waves were observed, the calculated growth rate was small. This experiment directly confirms the generation of the electrostatic waves at the appropriate frequencies via Z-mode cyclotron maser in the auroral ionosphere.

The rocket experiments did not detect the mode-converted electromagnetic radiation. The HIBAR data thereby place an upper bound of about 3% on the mode conversion efficiency in the observed example. However, based on the observed upper hybrid wave intensities, mode conversion efficiency of only 1% is required to explain the intensity of the emissions observed at ground level. Future rocket experiments will include higher dynamic-range receivers in order to directly detect the mode converted waves and estimate the mode conversion efficiency. Little is known about the mode conversion mechanism. However, the auroral ionosphere is characterized by sharp density gradients which could result in refraction and direct linear conversion between electrostatic Langmuir-upper hybrid waves and LO-mode electromagnetic waves via the so-called Ellis window (see for example Yoon et al., 1998b). The auroral ionosphere also is highly irregular on many spatial scales, the condition for linear mode conversion via interaction with low frequency density irregularities as described by Bell and Ngo (1990). Nonlinear mode conversion mechanisms may also play a role.

For waves just below the matching condition $f_{uh} \approx 2f_{ce}$ the refractive index surface implies trapping in electron density enhancements. Both the Porcupine [Carlson et al., unpublished manuscript, 1987] and the HI-

BAR (Samara et al., 2004) rocket experiments show evidence for such wave trapping. Yoon et al. (2000) show that if the cross-sectional dimension of a field-aligned density enhancement is less than about 1 km, the upper hybrid waves generated at the matching conditions in such an inhomogeneous plasma would be discretized. Essentially, the wave trapping results in standing waves at certain selected eigenfrequencies for which the wave phase integrated between reflection points equals an odd-integer number of $\pi/2$. Yoon et al. (2000) calculate these eigenmodes using WKB approach and find two characteristic frequency gaps associated with the eigenmode structure, a relatively wide gap associated with the radial mode structure which might be of order 1 kHz for an 800-m density enhancement, for example, and a relatively narrower gap that might be of order 100 Hz for an 800-m density enhancement. The auroral ionosphere usually contains field-aligned irregularities of many different scales ranging from meters to kilometers. These irregularities might in some cases be generated by the very high frequency waves that are subsequently trapped in them, or they may more likely be pre-existing irregularities caused by electron impact ionization on small scales or filementation or other large scale instabilities such as $E \times B$ instability acting on density gradients. In any case, the conditions are often favorable to formation of upper hybrid waves with discrete frequency structure due to the inhomogeneous background plasma, and the mode converted electromagnetic waves would retain this frequency structure.

Figure 3 shows an example high-resolution spectrogram of the upper hybrid waves observed near the matching condition $f = f_{uh} = 2f_{ce}$ in the HIBAR rocket experiment (right panel) and a similar high-resolution spectrogram of auroral roar emissions observed at ground level. Strikingly both phenomena are characterized by fine structures. The auroral roar is composed of multiple discrete features separated by approximately 1 kHz and decreasing and increasing together in frequency (LaBelle et al., 1995; Shepherd et al., 1998). In Figure 3, as in many examples, two frequency separations are observed: a narrow band of hundreds of Hertz and a wider gap of several kHz. The rocket data show a similar structure for the upper hybrid waves. This structure matches qualitatively that predicted by the wave trapping theory of Yoon et al. (2000).

In summary, the auroral roar emission, first identified by Kellogg and Monson (1979), has become fairly well understood in broad outline. The evidence is pretty strong that Z-mode maser mechanism acting near the matching conditions $f = f_{uh} = 2f_{ce}$ can generate electrostatic waves with ~ 10 mV/m amplitudes in the F-region when an auroral electron beam is present. There is evidence that such waves are trapped in density enhancements as theory says should happen. There is also evidence, both in the direct measurements of upper hybrid waves and in the ground-level measurements of auroral roar, for fine frequency structure which qualitatively matches that predicted by theory
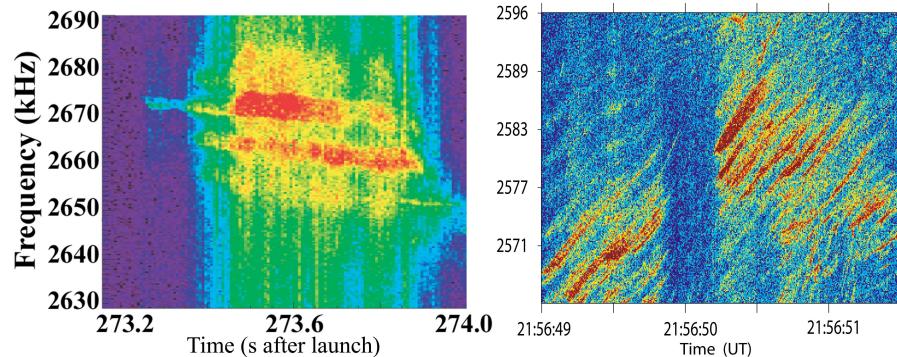
Figure 3. Fine structures characterizing upper hybrid waves and related electromagnetic auroral roar emissions: (a) Spectrogram of electric field fluctuations measured at an altitude of 277 km in the auroral ionosphere above Alaska, showing electrostatic upper hybrid waves with banded structure; (b) spectrogram of magnetic field fluctuations measured at ground level at South Pole, showing electromagnetic auroral roar with similar band structure.

of generation of discrete eigenmodes when the Z-mode maser acts in an inhomogeneous plasma characterized by field-aligned density irregularities with scale sizes of order tens to hundreds of meters. Quantitative comparison of the discrete eigenmode theory and the fine frequency structure observations is still lacking. The ingredient which is least well understood is the mode conversion mechanism, though a host of possibilities exist which could provide the required $\approx 1\%$ mode-conversion efficiency. There are a host of other interesting time- and frequency-variations characterizing auroral roar emissions which defy precise explanation but may result from propagation effects or modulations of the causative auroral electron beams (e.g., Hughes and LaBelle, 2001a). Understanding these phenomena is the focus of current research. Possibly the auroral roar may prove to be a valuable means of remotely sensing ionospheric plasma processes in the aurora.

2.2. AURORAL MF-BURST EMISSIONS

Figure 1 shows an example of an MF-burst emission occurring between 0652–0712 UT at 1.5–3.8 MHz. The emission thus coincides with the auroral substorm onset at 0652 UT; it is also coincident with impulsive auroral hiss at frequencies below 1 MHz, which is known to be associated with substorm onsets. The power spectral density of the MF-burst is in the range 10^{-13}–10^{-12} V^2/m^2Hz, comparable to that of auroral roar emission, but its bandwidth is ten times as great, implying that MF-burst requires an average source power of ~ 100 W assuming it originates from an isotropic source at

F-region altitudes. However, high time-resolution measurements of LaBelle et al. (1997) suggest that auroral MF-burst consists of impulses of submillisecond duration, far faster than resolved in the spectrogram of Figure 2, and in this case the peak power required of an isotropic source might be as high as 1 kW (less if the source is beamed, of course). Polarization measurements show that the principal component of MF-burst, at frequencies above 1.5 MHz, is polarized consistent with O-mode propagation in the ionosphere (Shepherd et al., 1997).

The generation mechanism of auroral MF-burst remains unknown and little studied. The O-mode polarization suggests that MF-burst may consist of mode-conversion radiation similar to auroral roar. The striking correlation between MF-burst and auroral hiss, clearly visible in Figure 1, suggests a close connection between the generation mechanisms of these two emissions. (In general, the correlation between MF-burst and auroral hiss is not one-to-one; MF-burst occasionally occurs without corresponding hiss, and auroral hiss more commonly occurs unaccompanied by MF-burst.) Examining the dispersion relation for normal modes of a homogeneous plasma at frequencies associated with electrons, two modes stand out for which the index of refraction greatly exceeds unity: The whistler mode on its resonance cone, and the upper-hybrid/slow Z-mode. Even the relatively low energy (1-10 keV) electron beam has particle velocities exceeding the phase velocities of these modes, and hence the auroral electrons could Cherenkov radiate into each of these modes, yielding auroral hiss in the case of the whistler mode, and, following mode conversion, auroral MF-burst in the case of the upper-hybrid/slow Z-mode. If the two waves extracted energy from different parts of the electron distribution function, they could exist simultaneously because they would not compete directly for the same energy source; for example, if the hiss were generated via Landau resonance with parallel electrons whereas the upper hybrid/slow Z-modes were generated via resonance with perpendicular electrons. In this case, however, the interaction would probably need to take place over a range of altitudes in order to explain the broad frequency range of each emission, so the auroral electrons would need to have sufficient free energy to excite waves over a range of altitudes. Simultaneous ground-based MF-burst and incoherent scatter radar observations in Sondrestrom, Greenland, provide evidence that the upper frequency of MF-burst emissions is connected to the maximum plasma or upper hybrid frequency in the overlying ionosphere, as would be predicted if MF-burst is emitted at the local plasma or upper hybrid frequency over a range of altitudes. Alternative generation mechanisms have appeared in the literature: LaBelle et al. (1997) suggested excitation of Langmuir waves over a range of altitudes on the bottomside of the F-region, followed by mode conversion, and Sotnikov et al. (1996) suggested excitation of electron acoustic waves covering a broad band of frequencies up to the local plasma frequency, followed by mode conver-

sion. Quantitative work on these MF-burst generation mechanisms is not yet sufficient to enable experimental tests.

On occasion MF-burst emissions show interesting frequency structure that may provide further hints to its origin. Weatherwax et al. (1994) reported a gap in the MF-burst spectrum near $2f_{ce}$, especially in the case that auroral roar emissions occur around the same time at that frequency. Figure 1 shows an example of this $2f_{ce}$ frequency gap. Occasionally, MF-burst extends below 1.5 MHz, which is approximately the electron gyrofrequency. In these cases, the emission spectrum exhibits a sharp lower cutoff at about 1.5 MHz and a gap before resuming below that frequency. (LaBelle and Treumann (2002) show an example.) This latter feature could indicate that MF-burst consists of two modes in the ionosphere: LO-mode above f_{ce} and whistler mode below f_{ce}. The polarization of the component below f_{ce} has never been measured to confirm this hypothesis.

Auroral MF-burst is the least-known of the types of spontaneous emission from aurora. The first unambiguous ground-level observations of MF-burst occurred relatively recently (Weatherwax et al., 1994), although some earlier observations may have been related to MF-burst (for example, bursty emissions reported by Parthasarathy and Berkey (1964) and Kellogg and Monson (1979).) MF-burst emissions typically last only a few minutes at substorm onset, and this short duration makes it impossible to plan a rocket launch into MF-burst. (The example in Figure 1 which lasts 20 minutes is unusual in this regard.) Recently, Burchill and Pfaff (2005) reported broadband impulsive emissions in the MF-burst frequency range detected by FAST satellite wave receivers during overflights of the auroral zone; if associated with MF-burst, these observations and especially the simultaneous particle measurements available in the FAST data set may provide a breakthrough in understanding the MF-burst generation mechanism.

2.3. Auroral Hiss

The third type of spontaneous emission from the auroral ionosphere is auroral hiss. Auroral hiss was first reported over 70 years ago (Burton and Boardman, 1933), although it is controversial whether the initial observation was really auroral hiss rather than lightning-generated atmospherics (Morgan, 1977). Over this long length of time, a huge literature has developed on auroral hiss, and thorough reviews have been written by Makita (1979), Sazhin et al. (1993), and LaBelle and Treumann (2002). Ground-level observations of auroral hiss fall into two categories, impulsive hiss and continuous hiss. Continuous auroral hiss is long-duration but usually limited to VLF (below about 30 kHz). Impulsive auroral hiss usually lasts only a few minutes at substorm onsets but extends up as high as 1 MHz.

Figure 1 shows an example of impulsive auroral hiss occurring between 0652 and 0712 UT at frequencies up to 800 kHz. The coincidence with auroral MF-burst is stiking in this example as noted above. As with the MF-burst, the impulsive auroral hiss is associated with substorm onset which in this example occurs at 0652 UT. The power spectral density of impulsive auroral hiss typically peaks at VLF and decreases monotonically with frequency until the emission blends into the noise level at ≤ 1 MHz. The power spectral density at LF is of order 10^{-12} v^2/m^{2}Hz, comparable to that of auroral roar and MF-burst. On some occasions, only the LF component of auroral hiss is detected (Morgan, 1977; LaBelle et al., 1998).

As noted above, the whistler mode is characterized by a resonance cone for which the index of refraction greatly exceeds unity. Hence the non-relativistic 1-10 keV auroral electrons can resonate with these waves. The generally accepted generation mechanism for auroral hiss is coherent amplification of the whistler mode waves on the resonance cone via Landau resonance with the auroral electrons. Maggs (1976) presented the first detailed calculations of auroral hiss spectra from this mechanism. Convection of the waves away from the resonance condition with the auroral electrons significantly affects their amplification, and therefore the geometry and other characteristics of the auroral arc play a big role in determining the auroral hiss spectrum. At low frequencies, the ray direction of whistler modes on the resonance cone is roughly parallel to the magnetic field, so the waves move fairly rapidly out of resonance with the electron beam; at high frequencies, the ray direction of the resonance cone waves is roughly perpendicular to the magnetic field, so the waves spend a longer time at a given altitude but they may move out of the auroral arc which is limited in extent, especially in the north-south direction. Strong refraction of these waves out of the auroral arc limits their growth. Combining these effects in a realistic auroral arc model, Maggs (1976, 1978) predicts auroral hiss spectra that peak at VLF and decrease monotonically with frequency at LF. Furthermore, the power spectral densities predicted by the theory are comparable to those observed with satellites and rockets and, accounting for attenuation, at ground-level. Because of this striking quantitative agreement, the convective beam amplification (CBA) mechanism has become broadly accepted as the source of auroral hiss. Subsequent rocket measurements have confirmed the theory in some cases (e.g., Bering et al., 1987).

Despite the success of the CBA model, several features of auroral hiss remain mysterious. For example, propagation of auroral hiss from ionospheric sources to ground-level has never been fully explained. In order to reach ground-level, the waves must convert from whistler mode to free-space mode at the bottom of the ionosphere (called the Earth-ionosphere boundary). At VLF, this boundary may be treated as sharp. Because the whistler mode index of refraction greatly exceeds unity while that of the free-space mode equals

unity, the waves in effect impinge on the boundary from the high index of refraction side, so unless their wave vector is nearly perpendicular to the boundary, or vertical in this case, they are totally internally reflected and do not reach ground-level. The problem is, it is difficult to generate waves on the whistler mode resonance cone at ionopsheric altitudes and have them propagate to the Earth-ionosphere boundary with their wave-vector perpendicular. Matsuo et al. (1998) calculate that such propagation cannot happen in a slab mode ionosphere but claim that it can occur if horizontal gradients are included such as field-aligned ducts or a plasmapause gradient. However, Sonwalkar and Harikumar (2000) argue that Matsuo et al. (1998) achieve this result by starting the waves with their wave-vector unrealistically far off of the resonance cone. They suggest that if the waves start sufficiently close to the resonance cone to be excited by 1-10 keV auroral electrons, those waves cannot be sufficiently refracted to reach the Earth-ionosphere boundary with vertical wave vector, even if realistic horizontal gradients are included. Sonwalkar and Harikumar (2000) put forward a model whereby the whistler waves are scattered from much smaller scale density gradients such as those associated with meter-scale irregularities which are abundant in the auroral ionosphere. Through such scattering the whistler modes near the resonance cone can be promptly converted to whistler modes well off of the resonance cone, including waves that can reach the Earth-ionosphere boundary with near-vertical wave vectors. Interestingly, despite ground-level observations of auroral hiss going back 70 years, how the auroral hiss reaches ground level remains a controversial issue.

Figure 4 shows examples of high resolution spectra of auroral waves. The top panel represents ground-level measurements of magnetic field at frequencies up to 200 kHz recorded at South Pole, Antarctica; 5 s of data are shown. The bottom two panels represent rocket-borne measurements of electric field from two different experiments launched from Poker Flat, Alaska. The ground-based data (top panel) show the typical LF auroral hiss which consists of signals whose power spectral density peaks at VLF, below the receiver's range, and extends into the LF range with power spectral density decreasing with frequency. The spectrum is broadband and featureless as is considered typical for auroral hiss, which is so-named because the VLF (audio frequency) component usually sounds like a featureless "hiss" when played through a loudspeaker. This broadband hiss is well-described by calculations using CBA theory (e.g., Maggs, 1976, 1978). However, superposed on the broadband hiss in Figure 4 are discrete features which stand out from the background hiss. These features consist of multiple spectral lines with bandwidth of order 10 kHz superficially resembling the fine structure of auroral roar shown in Figure 3. Remarkably, these 20–40 kHz features stand out despite the high-pass filter which severely attenuates the background broadband auroral hiss. The power spectral density of these discrete features

exceeds that of the broadband auroral hiss by up to 40 dB, although its total power is comparable since its bandwidth is smaller by about the same factor. Similar fine structure of auroral hiss has been reported at VLF (Ungstrup, 1971), and at both VLF and LF other types of fine structures have been reported in ground-level data (e.g., Siren, 1975; Ye and LaBelle, 2005). So far no theory predicts such highly peaked spectra with such high power spectral density, though it cannot be excluded that the CBA theory might predict such spectra for the right combination of plausible auroral arc geometry and electron distribution functions.

The bottom panels of Figure 4 show two types of structured whistler mode waves observed at rocket altitudes. The middle panel shows two seconds of data from the SIERRA rocket experiment, launched January 14, 2002, from Alaska. The spectrogram spans 250–650 kHz. Broadband auroral hiss appears intermittently, extending from below 250 kHz to about 450 kHz. However, imbedded within it are dramatic fine structures which dominate the spectrogram. These include "stripes": multiple narrow-band features which decrease in frequency from about 500 kHz to about 250 kHz during about 0.5 seconds, and which have power spectral density comparable to the background broadband hiss but bandwidth of order 10 kHz (Samara and LaBelle, unpublished manuscript, 2006). More intense, impulsive patchy features occur at 370–570 kHz with bandwidth of order 100 kHz and power spectral density 10–20 dB higher than the broadband auroral hiss or the stripes. The bottom panel of Figure 4 shows five seconds of data from the PHAZE-II rocket experiment, launched February 10, 1997, from Alaska. The spectrogram spans 200–600 kHz and shows a different type of whistler mode fine structure which occurs when $f_{pe} < f_{ce}$: long-duration, constant-frequency "bands" with 10–20 kHz bandwidth and millivolt per meter amplitudes (McAdams et al., 1999). These signals last up to 30 s of the rocket flight, corresponding to about 30 km of altitude, and they are often punctuated by intense Langmuir waves where their frequencies match the local plasma frequency. McAdams et al. (1999) interpret them as signatures of intermittent intense Langmuir waves which mode convert to whistler mode in the region where $f_{pe} < f_{ce}$. Previous lower-resolution rocket and satellite experiments detected hints of this zoo of fine structures within auroral hiss (e.g., LaBelle et al., 1999; Shutte et al., 1997; Beghin et al., 1989). The frequency range and some of the characteristics of these fine structures correspond to those observed at much higher altitudes in Auroral Kilometric Radiation (AKR), and Oya et al. (1985) put forth that AKR, which normally propagates upward from sources at ~ 5000 kilometers altitude, can under some conditions couple to the whistler mode and propagate to ionospheric altitudes. If so, such "leaked AKR" could explain these fine structure observations and possibly some of those detected at ground level. However, this hypothesis has not been confirmed by other theoretical or observational studies and remains highly controversial.
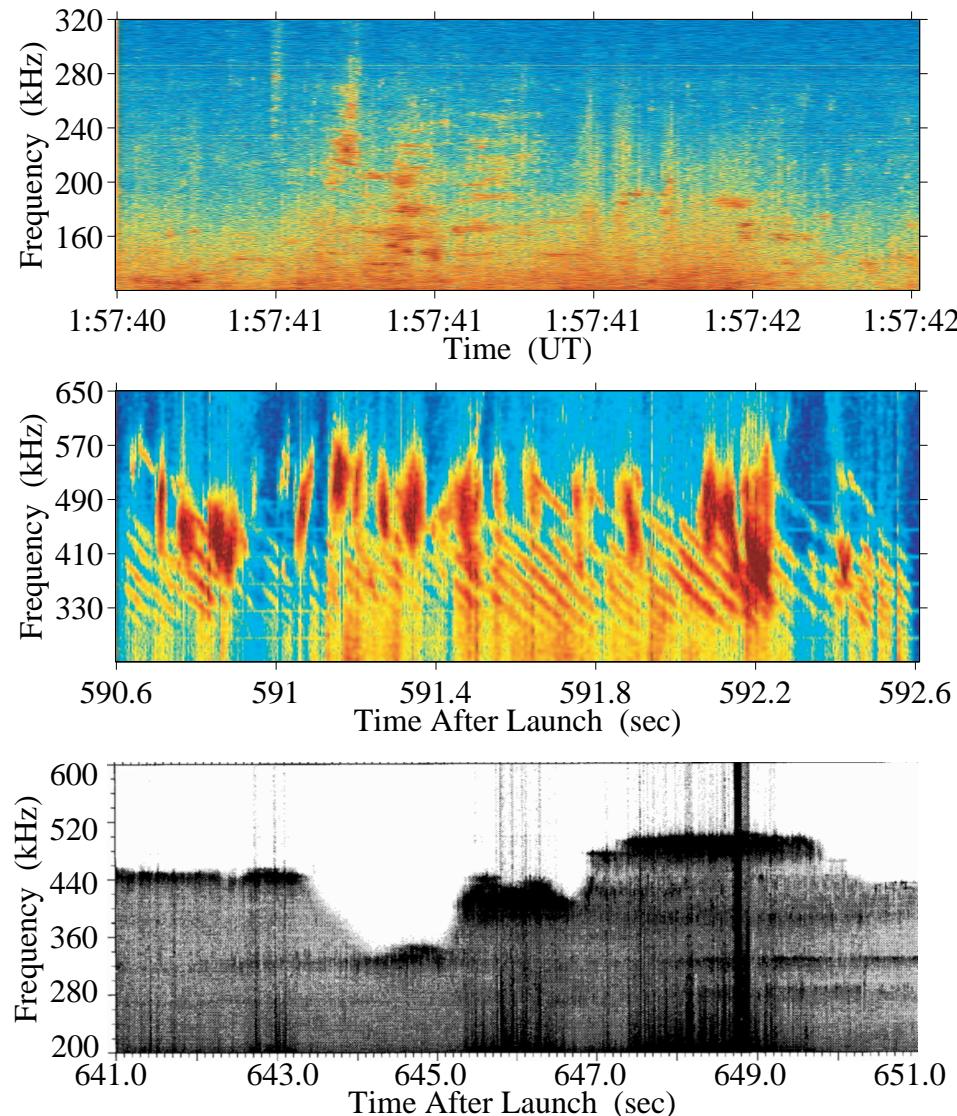
Figure 4. Fine structures characterizing whistler mode waves in the auroral ionosphere: (a) 120–320 kHz spectrogram of magnetic field fluctuations measured at ground-level at South Pole (2 s of data shown); (b) 250–650 kHz spectrogram of electric field fluctuations measured with the SIERRA rocket in Alaska (2 s of data shown); and (c) 200-600 kHz spectrogram of electric field fluctuations measured with the PHAZE-II rocket in Alaska (5 s of data shown). Different types of discrete features occur imbedded within the usual broadband auroral hiss.

In summary, impulsive auroral hiss, which occurs at substorm onset, often consists of a broadband spectrum of waves extending from VLF up to 1 MHz. Both the spectral shape and amplitude of this broadband hiss is well predicted by the CBA mechanism using reasonable models of the auroral electron distribution function and auroral arc. However, how these signals reach ground level remains controversial. At times various types of discrete whistler mode signals are imbedded in the broadband background auroral hiss, sometimes with power spectral densities greatly exceeding those of the background hiss. The origin of these signals is controversial. Despite 70 years of observations, auroral hiss is far from fully understood. Far more details of auroral hiss observations and theory can be found in comprehensive reviews (e.g., Sazhin et al., 1993; LaBelle and Treumann, 2002).

3. Summary

The discussion above focusses on three types of spontaneous emissions from the auroral ionosphere at radio frequencies: Auroral hiss, MF-burst, and roar, all of which can be observed at ground-level. Ground-level auroral roar occurs continuously for long durations, sometimes hours, preceding auroral substorms but becomes intermittent or absent after substorm onset, presumably due to poor propagation conditions which prevent the emission from reaching ground level. Impulsive auroral hiss and MF-burst are observed only for a few minutes at substorm onsets, presumably because the source conditions operate then but not at other times, and despite the relatively poor propagation conditions characterizing substorm onset. Auroral roar contains fine frequency structures which may carry information about ionospheric density irregularities. Auroral hiss and MF-burst may prove useful for timing and location of substorm onsets. Auroral hiss includes fine structures whose origin is not understood.

These three emissions are all relatively low-power in comparison with Auroral Kilometric Radiation (AKR). However, AKR cannot be considered an ionospheric emission because it is generated well above the ionosphere at thousands of kilometers altitude and beamed upward away from the ionosphere. Nevertheless, there are similarities between AKR and the lower altitude ionospheric radio emissions. The mechanism for most AKR is believed to be the X-mode cyclotron maser mechanism acting where the AKR frequency matches the local electron gyrofrequency, at altitudes of several thousand kilometers. This mechanism is similar to the Z-mode maser mechanism put forth to explain auroral roar emissions at F-region altitudes, except that in the AKR case the electrons directly excite the escaping X-mode and the radiated power is enormously greater. AKR exhibits fine structures which superficially resemble those which characterize auroral roar, but the mechanism

is unlikely to be the same since the modes are different. Calvert (1982) put forth a model for AKR fine structure that involved a wave feedback path with the wave path being an integer multiple of wavelengths, caused for example by reflection of the X-mode waves generated by the maser mechanism from reflection surfaces which converge with altitude (to compensate for refraction of the X-mode waves within the source region). Analogous to an optical laser, as long as the net gain around the feedback loop exceeds unity, the waves would grow to saturation with a tiny bandwidth, essentially monochromatic. In this model, many such compact radiators provide the source of the many narrow-band fine structures composing AKR. Observations of extremely narrow bandwidth AKR structures, of order 1 Hz, support this mechanism (Baumback and Calvert, 1987), but direct evidence is lacking. Recently, Mutel et al. (2005) propose an alternative model for one particular type of AKR fine structure, striated AKR, which consists of multiple descending tones. They put forth that the electric fields associated with ion phase space holes, features known to exist in the auroral acceleration region where AKR originates, locally distort the electron distribution function in such a way as to modulate the emission of AKR via the X-mode maser mechanism. Quantitative modelling shows that the modulation and frequency slope predicted by the theory agree quantitatively with those observed in striated AKR. Recently Bhesso et al. (2006) and Menietti et al. (2006) have explored variations of this model involving electromagnetic ion cyclotron waves rather than ion holes as the causative agent. Over the years, other mechanisms have been put forward to explain AKR fine structure (e.g., Louarn and Le Quéau, 1996a,b; Yoon and Weatherwax, 1998), but despite the success of certain models explaining subsets of the AKR fine structure, it remains an open problem.

The spontaneous auroral emissions also bear close similarities to stimulated electromagnetic emissions (SEE) generated by injection of high-power radio waves into the ionosphere (Thidé et al., 1982). A zoo of such emissions has been identified associated with different pump frequencies and powers (see review by Leyser, 2001). Some of these stimulated emissions closely resemble the natural spontaneous auroral emissions; in the one case the energy source is the injected pump wave, and in the other case, the auroral electrons.

The interconnections to other phenomena such as SEE and AKR provide a strong motivation for further investigation of the spontaneous emissions from the auroral ionosphere: not only are these an interesting component of Earth's radio environment which potentially play a significant role in that environment by interacting with particle populations and with other wave modes and which may serve as a useful ground-based diagnostic of ionospheric plasma processes, but understanding these auroral emissions originating a few hundred kilometers from Earth aids understanding of stimulated and spontaneous radio emission processes in a variety of other plasmas including other plane-

tary ionospheres and the solar wind. In some of these other cases, inaccessible to rocket and satellite experiments, the radio emissions are the primary means of sensing remote plasma processes.

Acknowledgements

References

Bale, S. D.: 1999, 'Observation of the topside ionospheric MF/HF radio emission from space'. *Geophys. Res. Lett.* **26**, 667.

Baumback, M. M. and W. Calvert: 1987, 'The minimum bandwidths of auroral kilometric radiation'. *Geophys. Res. Lett.* **14**, 119.

Beghin, C., J. L. Rauch, and J. M. Bosqued: 1989, 'Electrostatic Plasma Waves and HF Auroral Hiss Generated at Low Altitude'. *J. Geophys. Res.* **94**, 1359.

Bell, T. F. and H. D. Ngo: 1990, 'Electrostatic lower hybrid waves excited by electromagnetic whistler mode waves scattering from planar magnetic field aligned plasma density irregularities'. *J. Geophys. Res.* **95**, 149.

Benson, R. F. and H. K. Wong: 1987, 'Low-altitude ISIS 1 observations of auroral radio emissions and their significance to the cyclotron maser instability'. *J. Geophys. Res.* **92**, 1218.

Bering, E. A., J. E. Maggs, and H. R. Anderson: 1987, 'The Plasma Wave Environment of and Auroral Arc'. *J. Geophys. Res.* **92**, 7581.

Bhesso, N., D. Menietti, and A. Bhattacharjee: 2006, 'PIC Simulations of the Generation of Striated AKR'. *EOS Trans. Am. Geophys. Union* **87**(36), SM33B–03.

Burchill, J. K. and R. F. Pfaff: 2005, 'FAST observations of auroral MF emissions and precipitating electrons'. *EOS Trans. Am. Geophys. Union* **86**(52), SM44A–01.

Burton, E. T. and E. M. Boardman: 1933, 'Audio-frequency atmospherics'. *Proc. Inst. Radio Eng.* **21**, 1476.

Calvert, W.: 1982, 'A feedback model for the source of auroral kilometric radiation'. *J. Geophys. Res.* **87**, 8199.

Chiu, Y. T. and M. Schulz: 1978, 'Self-consistent particle and parallel electrostatic field distributions in the magnetospheric-ionospheric auroral zone'. *J. Geophys. Res.* **83**, 629.

Gough, M. P. and A. Urban: 1983, 'Auroral Beam/Plasma Interaction Observed Directly'. *Planet. Space Sci.* **31**, 875.

Hughes, J. M. and J. LaBelle: 1998, 'The latitude dependence of auroral roar'. *J. Geophys. Res.* **103**, 14911.

Hughes, J. M. and J. LaBelle: 2001a, 'First observations of flickering auroral roar'. *Geophys. Res. Lett.* **28**, 123.

Hughes, J. M. and J. LaBelle: 2001b, 'Plasma conditions in auroral roar source regions inferred from radio and radar observations'. *J. Geophys. Res.* **106**, 21157.

James, H. G., E. L. Hagg, and L. P. Strange: 1974, 'Narrowband radio noise in the topside ionosphere'. *AGARD Conf. Proc.* **AGARD-CP-138**, 24–1–24–7.

Kaufmann, R. L.: 1980, 'Electrostatic wave growth: secondary peaks in measured auroral electron distribution functions'. *J. Geophys. Res.* **85**, 1713.

Kellogg, P. J. and S. J. Monson: 1979, 'Radio emissions from the aurora'. *Geophys. Res. Lett.* **6**, 297–300.

LaBelle, J., K. L. McAdams, and M. L. Trimpi: 1999, 'High-frequency and time resolution rocket observations of structured low- and medium-frequency whistler mode emissions in the auroral ionosphere'. *J. Geophys. Res.* **104**, 28,101.

LaBelle, J., S. G. Shepherd, and M. L. Trimpi: 1997, 'Observations of auroral medium frequency bursts'. *J. Geophys. Res.* **102**, 22,221.

LaBelle, J. and R. A. Treumann: 2002, 'Auroral Radio Emissions, 1. Hisses, Roars, and Bursts'. *Space Sci. Rev.* **99**, 295–440.

LaBelle, J., M. L. Trimpi, R. Brittain, and A. T. Weatherwax: 1995, 'Fine structure of auroral roar emissions'. *J. Geophys. Res.* **100**, 21953.

LaBelle, J., A. T. Weatherwax, J. Perring, E. Walsh, M. L. Trimpi, and U. Inan: 1998, 'Low-frequency impulsive auroral hiss observations at high geomagnetic latitudes'. *J. Geophys. Res.* **103**, 20459–20468.

Leyser, T. B.: 2001, 'Stimulated electromagnetic emissions by high-frequency electromagnetic pumping of the ionospheric plasma'. *Space Sci. Rev.* **98**, 223 – 328.

Louarn, P. and D. Le Qu´eau: 1996a, 'Generation of auroral kilometric radiation in plasma cavities, I, Experimental study'. *Planet. Space Sci.* **44**, 199.

Louarn, P. and D. Le Qu´eau: 1996b, 'Generation of auroral kilometric radiation in plasma cavities, II, The cyclotron maser instability in small size sources'. *Planet. Space Sci.* **44**, 211.

Maggs, J. E.: 1976, 'Coherent Generation of VLF Hiss'. *J. Geophys. Res.* **81**, 1707.

Maggs, J. E.: 1978, 'Electrostatic Noise Generated by the Auroral Electron Beam'. *J. Geophys. Res.* **83**, 3173.

Makita, K.: 1979, 'VLF/LF hiss emissions associated with aurora'. *Mem. Nat. Int. Polar Res.* **16**, 1.

Matsuo, T., T. Nishiyama, and D. Matuhara: 1998, 'Propagation of a quasi electrostatic whistler mode auroral hiss to the ground'. *Proc. NIPR Symp. Upper Atmos. Phys.* **12**, 12–29.

McAdams, K. L., J. LaBelle, M. L. Trimpi, P. M. Kintner, and R. A. Arnoldy: 1999, 'Rocket observations of banded stucture in waves near the Langmuir frequency in the auroral ionosphere'. *J. Geophys. Res.* **104**, 28109.

Menietti, J. D., R. L. Mutel, O. Santolik, J. D. Scudder, I. W. Christopher, and J. M. Cook: 2006, 'Striated drifting auroral kilometric radiation bursts: Possible stimulation by upward traveling EMIC waves'. *J. Geophys. Res.* **111**, A04214, doi:10.1029/2005JA011339.

Morgan, M. G.: 1977, 'Wide-Band Observations of LF Hiss at Frobisher Bay ($L = 14.6$)'. *J. Geophys. Res.* **82**, 2377.

Mutel, R. L., D. Menietti, and I. Christopher: 2005, 'Striated AKR: A Remote Tracer of Solitary Ion Structures in the Magnetosphere'. *EOS Trans. Am. Geophys. Union* **86**(52), SM41B–1189.

Oya, H., A. Morioka, and T. Obara: 1985, 'Leaked AKR and Terrestrial Hectometric Radiations Discovered by the Plasma Wave and Planetary Plasma Sounder Experiments on Board the Ohzora (EXOS-C) Satellite- Instrumentation and Observation Results of Plasma Wave Phenomena'. *Geomag. Geoelectr.* **37**, 237.

Parthasarathy, R. and F. T. Berkey: 1964, 'Radio noise from the auroral electrons-I'. *J. Atmos. Terr. Phys.* **26**, 199.

Samara, M., J. LaBelle, C. A. Kletzing, and S. R. Bounds: 2004, 'Rocket observations of structured upper hybrid waves at $f_{uh} = 2f_{ce}$'. *Geophys. Res. Lett.* **31**(L22804), doi:10.1029/2004GL021043.

Sazhin, S. S., K. Bullough, and M. Hayakawa: 1993, 'Auroral hiss: A review'. *Planet. Space Sci.* **41**, 153–166.

Shepherd, S. G., J. LaBelle, and M. L. Trimpi: 1997, 'The polarization of auroral radio emissions'. *Geophys. Res. Lett.* **24**, 3161.

Shepherd, S. G., J. LaBelle, and M. L. Trimpi: 1998, 'Further investigation of auroral roar fine structure'. *J. Geophys. Res.* **103**, 2219.

Shutte, N., I. Prutensky, S. Pulinets, Z. Klos, and H. Rothkaehl: 1997, 'The charged-particle fluxes at auroral and polar latitudes and related low-frequency auroral kilometric radiation-type and high-frequency wideband emission'. *J. Geophys. Res.* **102**, 2105.

Siren, J. C.: 1975, 'Fast Hisslers in Substorms'. *J. Geophys. Res.* **80**, 93.

Sonwalkar, V. S. and J. Harikumar: 2000, 'An explanation of ground observations of auroral hiss: Role of density depletions and meter scale irregularities'. *J. Geophys. Res.* **105**, 18867–18883.

Sotnikov, V. I., D. Schriver, M. Ashour-Abdalla, and J. LaBelle: 1996, 'Generation of auroral radio waves by a gyrating electron beam'. *EOS Trans. Am. Geophys. Union* **77**, F544.

Thid´e, B., H. Kopka, and P. Stubbe: 1982, 'Observations of Stimulated Scattering of a Strong High Frequency Radio Wave in the Ionosphere'. *Phys. Rev. Lett.* **49**, 1561–1564.

Ungstrup, E.: 1971, 'Rocket Observation of VLF Hiss in Aurora'. *Planet. Space Sci.* **19**, 1475.

Weatherwax, A. T., J. LaBelle, and M. L. Trimpi: 1994, 'A new type of auroral radio emission observed at medium frequencies($\sim$ $1350 - 3700$ kHz) using ground-based receivers'. *Geophys. Res. Lett.* **21**, 2753.

Weatherwax, A. T., J. LaBelle, M. L. Trimpi, and R. Brittain: 1993, 'Ground-based observations of radio emissions near $2f_{ce}$ and $3f_{ce}$ in the auroral zone'. *Geophys. Res. Lett.* **20**, 1447–1450.

Weatherwax, A. T., J. LaBelle, M. L. Trimpi, R. A. Treumann, J. Minow, and C. Deehr: 1995, 'Statistical and case studies of radio emissions observed near $2f_{ce}$ and $3f_{ce}$ in the auroral zone'. *J. Geophys. Res.* **100**, 7745.

Ye, S. and J. LaBelle: 2005, 'High Resolution Measurement of LF Auroral Hiss at South Pole'. *EOS Trans. Am. Geophys. Union* **86**(18), SM43C–07.

Yoon, P. H. and A. T. Weatherwax: 1998, 'Theory for AKR fine frequency structure'. *Geophys. Res. Lett.* **25**, 4461–4464.

Yoon, P. H., A. T. Weatherwax, and J. LaBelle: 2000, 'Discrete electrostatic eigenmodes associated with ionospheric density structure: Generation of auroral roar fine frequency structure'. *J. Geophys. Res.* **105**, 27589.

Yoon, P. H., A. T. Weatherwax, and T. J. Rosenberg: 1996, 'Lower ionospheric cyclotron maser theory: A possible source of 2fce and 3fce auroral radio emissions'. *J. Geophys. Res.* **101**, 27,015.

Yoon, P. H., A. T. Weatherwax, and T. J. Rosenberg: 1998a, 'On the generation of auroral radio emissions at harmonics of the lower ionospheric electron cyclotron frequency: X, O and Z Mode maser calculations'. *J. Geophys. Res.* **103**, 4071.

Yoon, P. H., A. T. Weatherwax, T. J. Rosenberg, J. LaBelle, and S. G. Shepherd: 1998b, 'Propagation of medium frequency (1-4 MHz) auroral radio waves to the ground via the Z-Mode radio window'. *J. Geophys. Res.* **103**, 29267.

Chapter 5 – SPACE WEATHER

5.1 IONOSPHERIC EFFECTS OF SEVERE GEOMAGNETIC STORMS

(John C. Foster, MIT Haystack Observatory)

Abstract

Earth's ionosphere responds dramatically to severe geomagnetic storms. Coupled closely to the neutral thermosphere and the overlying magnetosphere, stormtime ionospheric effects appear with a wide variety of scale sizes, characteristics, time histories, and associated drivers. The strong interconnection between the magnetosphere and ionosphere points to the need to take a global perspective in characterizing and understanding this system. Penetrating electric fields and plasma redistribution combine to produce severe enhancements and reductions of ionospheric TEC (total electron content), steep localized density gradients, and stormtime space weather effects.

5.1.1 Introduction

Earth's ionosphere responds dramatically to severe geomagnetic storms. Coupled closely to the neutral thermosphere and the overlying magnetosphere, stormtime ionospheric effects appear with a wide variety of scale sizes, characteristics, time histories, and associated drivers. This report provides an overview of recent developments in observing, identifying, and understanding the causes of such phenomena. In keeping with the author's experience, the emphasis is placed meso-scale (100s – 1000s kilometer) ionospheric perturbations and the use of ground and space-based diagnostics to characterize them. As often occurs in a topical review, the work reported and illustrations used will reflect those areas most familiar to the author. Accordingly, emphasis is placed on recent observations of mid-latitude ionospheric disturbances and their relationship to the complex system of stormtime magnetosphere-ionosphere coupling.

Recent studies highlighting the disturbance effects observed in the ionosphere and magnetosphere during geomagnetic disturbances, and their space weather effects, have heightened interest in the processes occurring in the plasmasphere boundary layer (PBL) [Carpenter and Lemaire, 2004], the boundary between the corotating field lines of the plasmasphere and the outer magnetosphere. During storms, the effects of magnetospheric electric field and ExB plasma convection extend deep into the mid-latitude ionosphere producing dramatic effects. At the PBL, strong electric fields contribute to the formation of the deep mid-latitude density trough which spans the nightside while penetration electric fields uplift, destabilize, and perturb the low-latitude ionosphere. For example, Foster and Rich [1998] reported a sudden uplifting of the mid-latitude ionosphere, equatorward of the trough, during the early phase of strong storms due to the effects of prompt penetration electric fields. Destabilization of the F region at mid latitudes (up to ~35° invariant latitude) followed (cf. Figure 5-1).

Sub-Auroral Topside Perturbation

Millstone Hill Elevation Scan 21 Mar 90 02 UT
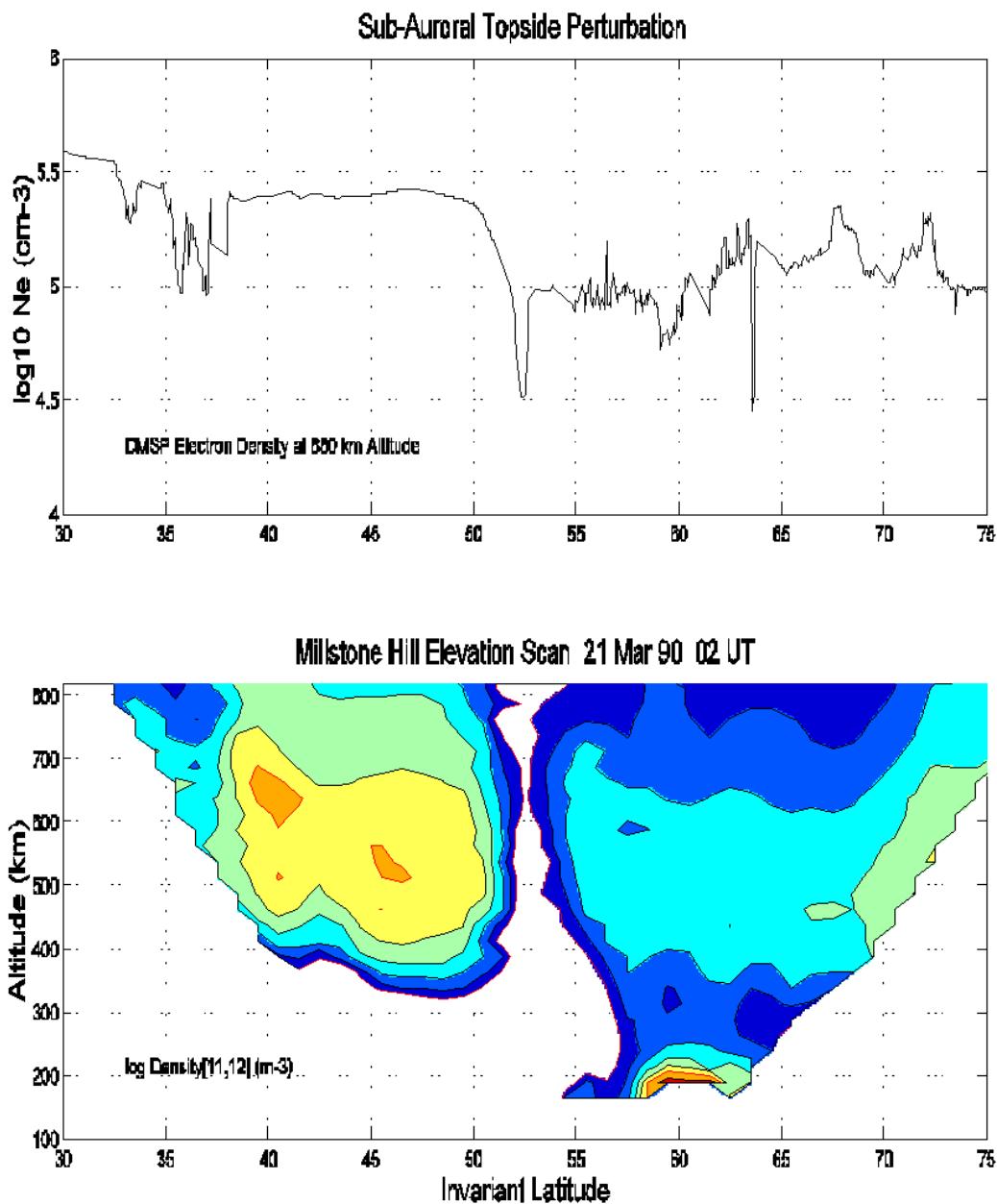
Figure 5-1: Millstone Hill ISR Elevation Scan (bottom) Identified a Deep Mid-Latitude Ionospheric Trough Near 53 deg Invariant Latitude at 21 MLT (02 UT) During the March 21, 1990 Storm. A pronounced low-latitude topside ionosphere density depletion was observed by both the radar and by the DMSP F-9 satellite more than 15° equatorward of the normal mid-latitude trough during the March, 1990 event. The destabilizing effects of a penetrating electric field were discussed by Foster et al. [1990].

5.1.2   Storm Positive Phase

Ionospheric storms are large global disturbances of the *F*-region electron density in response to geomagnetic storms. The terms positive and negative storm phase are used to describe large-scale increases (early storm phase) and decreases (late storm phase) in the ionospheric electron density regularly observed during geomagnetic storms. During the negative phase (usually beginning the second day of and event), the *F*-peak electron density in the mid-latitude ionosphere may be reduced by a factor of 2-5.

The magnetospheric energy input to the atmosphere at auroral latitudes is greatly increased during magnetic storms. Enhanced Joule heating at high latitudes reduces the normal poleward wind on the dayside and reinforces the regular equatorward wind on the nightside, creating a storm circulation that can transport air with increased molecular species to mid latitudes. As these neutral composition disturbances move to lower latitudes, the associated enhanced ion loss rate can result in significant decreases in the *F*-region electron density.

Large increases in the mid-latitude ionospheric *F*-region electron density and total electron content (TEC) are often observed in local dusk sector during magnetic storms. It was proposed in early investigations [Evans, 1973; Mendillo, 1973] that an uplifting of the F-layer by an eastward electric field and convergence in the east-west direction might be responsible for such dusk effect enhancements. Foster [1993] examined Millstone Hill incoherent scatter radar observations and suggested that the dusk effect during disturbed conditions can be associated with sunward convection (towards noon and poleward at ionospheric heights) of high-density plasma originating from lower latitudes. The resultant density enhancements at the equatorward edge of the dusk-sector ionospheric trough are termed storm-enhanced density (SED). The study of Buonsanto [1995] showed that both wind-induced uplift and advection from lower latitudes can be involved in the dusk-sector stormtime enhancements.

Huang et al [2005a] give a detailed review of the range of mechanisms which can give rise to positive phase mid and low-latitude density increases during storms. Foster et al. [2005a] and Yin et al., 2004] suggest that storm time TEC enhancements may corotate with the Earth once they are formed. However, some features observed by Huang et al [2005a] cannot be fully explained by a corotation process. During a particular storm being investigated, they noted that no positive phase enhancement occurred over the daytime US continent, although a pronounced enhancement was observed at east-coast longitudes. The driving eastward electric field lasted for many hours during the event and should have been effective in uplifting the ionosphere and continuously generating a dayside positive storm signature. The presence of positive phase in the Atlantic sector and its non-existence over the US continent implied that the associated processes might be more effective in generating a positive storm enhancement in the Atlantic sector. Similar observations of the longitudinal localization of stormtime electric field effects (with largest enhancement in the Atlantic sector) had been noted by Foster and Rich [1998]. Most recently, Foster and Coster [2006] investigated such a discrete localized enhancements of total electron content at dusk in the American sector. They found that the repeatability and geographic localization of this feature indicate a longitudinal specificity for the process associated with its formation.

5.1.3   Penetration Electric Fields

Under disturbed conditions the interplanetary electric field (IEF) can penetrate to the low- latitude ionosphere. Penetration of the IEF to the middle- and low-latitude ionosphere has been extensively studied through geomagnetic field perturbations and using the incoherent scatter radar chain at ~75° W geographic longitude which includes the Sondrestrom, Millstone Hill, Arecibo, and Jicamarca radars [Kelley et al., 1979; Gonzales et al., 1978; Buonsanto et al., 1999]. A shielding/overshielding mechanism has been proposed to explain how the interplanetary/magnetospheric electric field can, or cannot penetrate to the low-latitude ionosphere [Jaggi and Wolf, 1973]. The shielding of the inner magnetosphere/ionosphere from the high-latitude disturbance electric fields involves the magnetospheric ring current and current systems which interconnect the magnetosphere and ionosphere near the PBL. Penetration electric fields can uplift and perturb the low and mid-latitude ionosphere (cf. Figure 5-1) and lead to plasma redistribution (see Section 5.1.6 below) which circulates low-latitude ionospheric plasma to mid, high, and polar latitudes during strong storms(e.g. Foster et al [2005a;b]; Mannucci et al. [2005]).

Whereas a time scale for shielding/overshielding effects of ~30 min is usual, recent work by Huang et al. [2005b;2006] finds that when the IMF turns southward and remains stably southward for several hours, the dayside eastward ionospheric electric field at low latitudes is enhanced throughout the entire interval

of southward IMF. A similar enhancement of the westward ionospheric electric field is observed when the IMF turns northward and remains stably northward. That study indicates that the ionospheric electric field at low latitudes can be continuously enhanced over many hours during the main phase of magnetic storms. In sunlight, an eastward electric field will cause increases in the mid-latitude ionospheric electron density by moving the plasma particles upward into regions of decreased recombination loss, while additional ionization is produced at lower altitude. Decreases in the equatorial ionospheric electron density occur as the low-latitude plasma is redistributed poleward by an enhanced fountain effect.

5.1.4   Storm Enhanced Density (SED)

The overall stormtime enhancement of TEC at low and mid latitudes consists of two parts (cf. Figure 5-2 which presents a schematic representation.). First is the increase of TEC seen at and poleward of the crests of the equatorial anomalies associated with plasma uplift and redistribution from low to mid latitudes under the influence of the penetration electric field. This occurs inside the plasmasphere boundary layer and serves as an enhanced source population for the plumes of storm enhanced density (SED) [Foster, 1993] that are eroded from the lower-latitude ionosphere/ plasmasphere by disturbance electric fields.
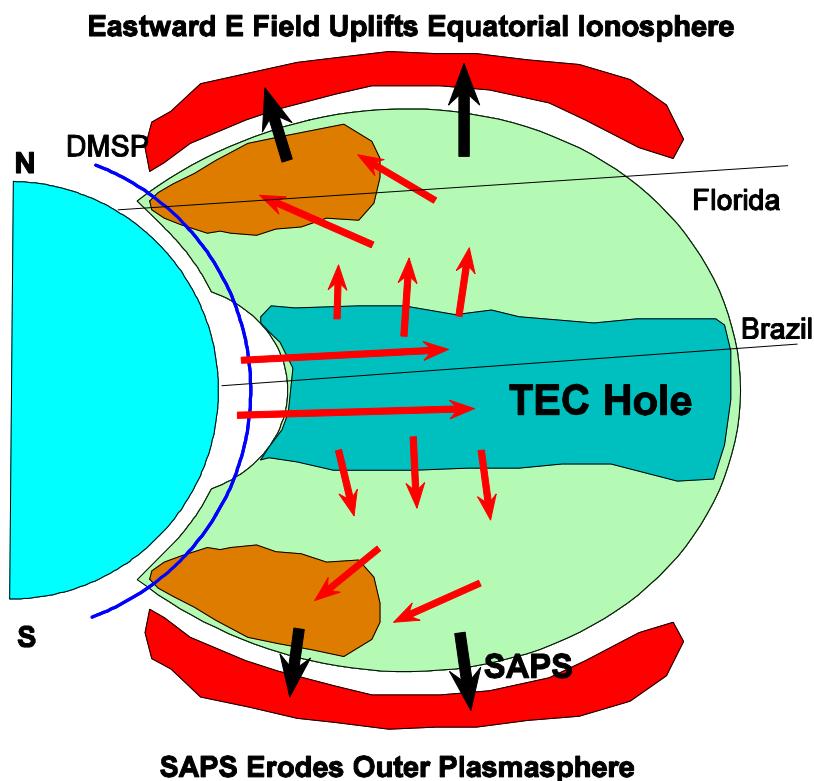
**Figure 5-2: The Dual Effects of Disturbance Electric Fields are Presented Schematically.
Undershielded penetration electric fields uplift the equatorial ionosphere
redistributing equatorial plasma poleward, while SAPS electric fields
strip away the enhanced outer layers of the plasmasphere.**

Enhanced cold plasma total electron content (TEC) is observed at mid latitudes, inside the PBL, in the post-noon beginning in the early phases of the storms, while Dst is falling (e.g. Buonsanto and Foster, 1993; Tsurutani et al., 2004). At the poleward extent of this region, elevated TEC is carried sunward and poleward in plumes of storm enhanced density (SED [Foster, 1993]). Figure 5-3 presents Millstone Hill incoherent scatter radar (ISR) elevation-scan observations of the latitude/altitude structure observed across

such a plume of storm enhanced density. F-region TEC (calculated for the ISR density-altitude profiles) approached 100 TECu within the SED plume and sharp TEC gradients (~50 TECu/deg latitude) were observed bordering the SED feature.
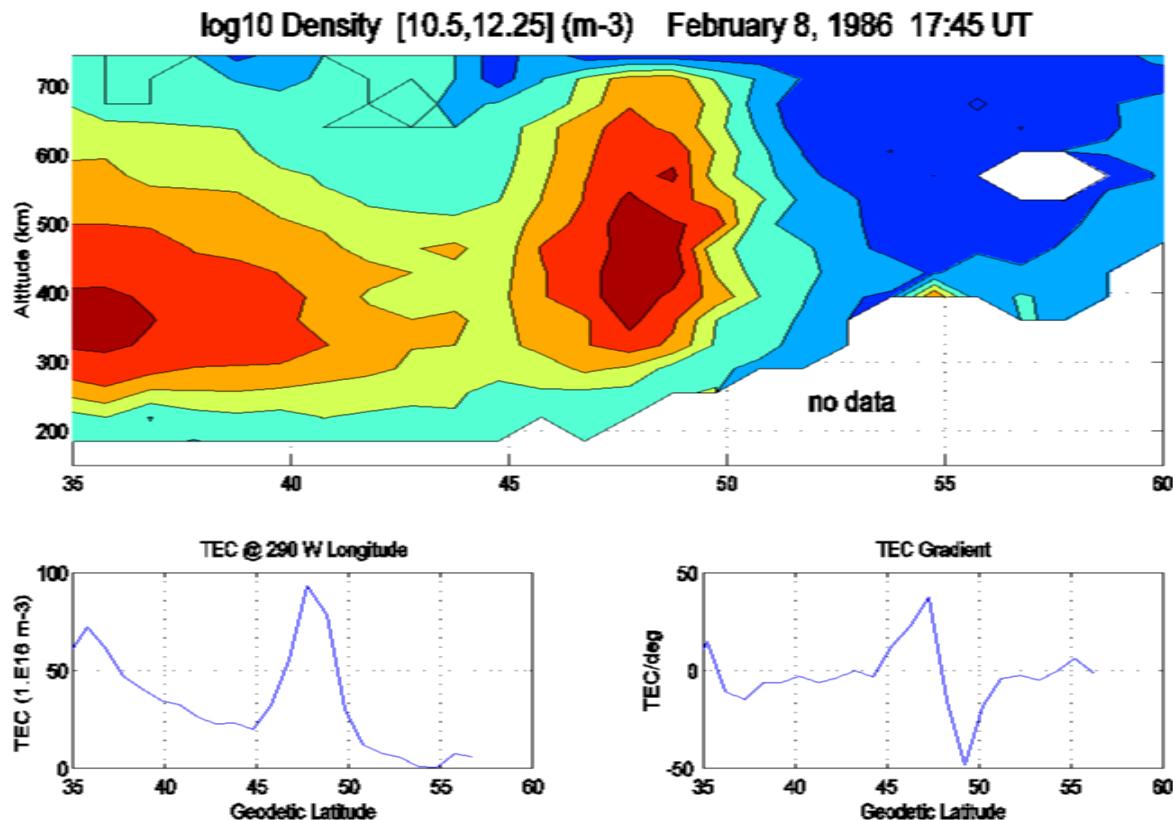
Figure 5-3: Iso-Density Contours Observed by the Millstone Hill Radar Scanning N-S Across a Region of Strong SED (Storm-Enhanced Density) Near Local Noon are Presented as a Function of Geodetic Latitude (Invariant Latitude = Geodetic Latitude + 11°). F-region TEC and TEC gradient across the region are derived from the radar elevation scan and are shown in the bottom panels. TEC near 48° N geodetic latitude is ~100 TEC units and the latitude gradient in TEC associated with the region of SED near 45° latitude was ~50 TEC/deg.

In the dusk sector where they are formed, the SED plumes are associated with the stripping away of the outer layers of the plasmasphere/ionosphere by the disturbance electric field. Storm enhanced density occurs when low-latitude cold plasma is transported sunward (towards noon) at the inner edge of the convection electric field. This process produces narrow plasmasphere drainage plumes (Sandel et al., 2001) of cold plasma which extend along magnetic field lines between the plasmasphere and the ionosphere (e.g. Chi et al., 2005). Figure 5-4 depicts the location and spatial extent of the SED features during a typical event. Ground-based GPS-TEC observations depict the SED plume spanning the North American continent from a source region over the US east coast. Mapping this feature into the magnetospheric equatorial plane (Figure 5-4 right) shows the relationship of the ionospheric SED plume to stormtime plasmasphere erosion. Combining ground (GPS TEC) and space-based (IMAGE EUV) thermal plasma imaging techniques, Foster et al. [2002] demonstrated that this ionospheric SED plume mapped into the low-altitude signature of a plasmaspheric drainage plume viewed from space. Cold plasma within the plume at both ionospheric and magnetospheric heights is streaming sunward towards the cusp (at low altitudes) and the dayside magnetopause (at high altitudes). Foster et al. [2004a] used direct radar observations of the sunward **ExB** plasma convection to quantify the flux of ions carried by the

SED plume to the noontime F-region in the vicinity of the cusp. Foster et al. [2005b] report observations which indicate that this streaming low-latitude plasma is also a source for the tongues of ionization which appear at polar latitudes during disturbed conditions (see Section 5.1.7 below).
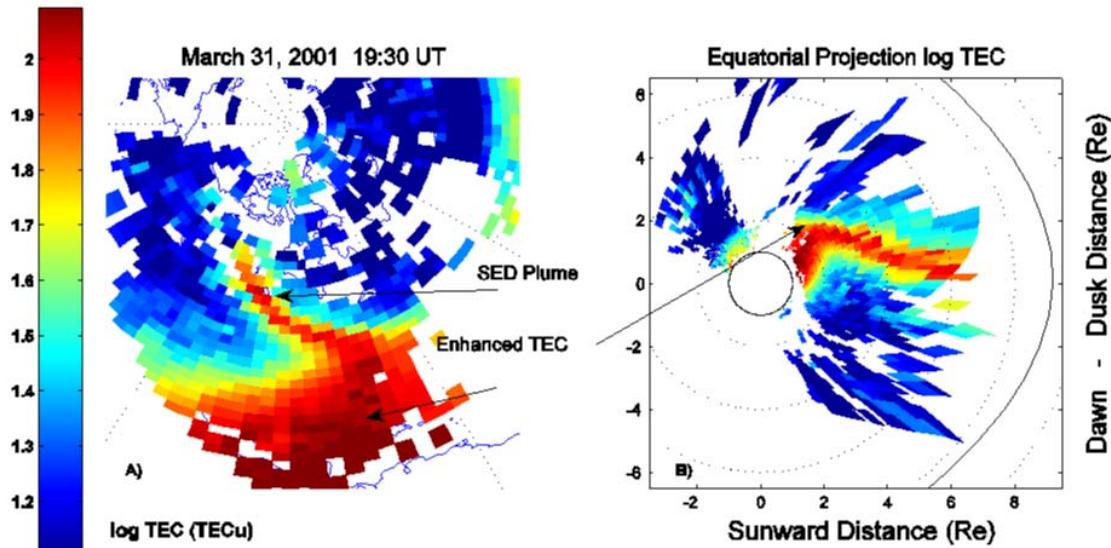
Figure 5-4: (A) A Region of Enhanced GPS TEC was Observed at the Base of the Plume of Storm Enhanced Density Seen over North America During the March 31, 2001 Event. (B) Projecting the GPS TEC Observations into the Magnetospheric Equatorial Plane Using Tysganenko Mapping (with the sun at the right), Indicates that the Enhancement at the Base of the Plume is Field Lines Threading the Outer Plasmasphere.

5.1.5   Sub-Auroral Polarization Stream (SAPS)

Electric fields are of prime importance in the formation and transport of stormtime ionospheric disturbances. The sub-auroral polarization stream (SAPS) electric field [Foster and Burke, 2002] refers to the region of enhanced poleward electric field which forms equatorward of electron precipitation in the dusk to midnight sector in disturbed conditions. SAPS is a magnetospheric electric field, whose characteristics are controlled by ionospheric conductance. SAPS forms as pressure gradients at the inner edge of the disturbance-enhanced ring current drive Region 2 field-aligned currents into the evening-sector ionosphere. Large poleward-directed electric fields are set up to drive closure currents across the low-conductivity region equatorward of the auroral electron precipitation. Observations of this phenomenon at ionospheric altitudes [Yeh et al., 1991; Foster and Burke, 2002; Foster and Vo, 2002] describe the occurrence characteristics and persistence of the SAPS electric field whose latitude extent spans the region between the electron plasma sheet and the outer reaches of the plasmasphere (the PBL). Significant magnetospheric effects occur as the SAPS electric field maps along field lines and is observed at higher altitudes (e.g. Rowland and Wygant [1998]).

Foster et al [2006] used magnetic field-aligned mapping between the ionosphere and the magnetosphere to intercompare ground-based observations of storm enhanced density (SED), and plasmasphere drainage plumes imaged from space by the IMAGE EUV imager, with the enhanced inner-magnetosphere/ ionosphere SAPS electric field which develops during large storms. They found that the inner edge of the SAPS electric field overlaps the erosion plume and that plume material is carried sunward in the SAPS overlap region. The two phenomena, SED in the ionosphere and the erosion plume at magnetospheric heights, define a common trajectory for sunward-propagating cold plasma fluxes in the midnight-dusk-postnoon sector. The SAPS channel at ionospheric heights and its projection into the equatorial plane

serve to define the sharp outer boundary of the erosion plume. The SAPS electric field abuts and overlaps both the plasmasphere boundary layer and the plasmasphere erosion plume from pre-midnight through post-noon local times.

5.1.6   Plasma Redistribution

While some ionospheric disturbances during geomagnetic storms are local effects, recent studies using ground-based ionospheric observations have found that the equatorial, mid-latitude, auroral, and polar regions are coupled by processes which redistribute cold ionospheric and plasmaspheric material throughout the interconnected system. In particular, cold plasma enhancements, created at equatorial and mid-latitudes by storm time processes, are carried poleward by disturbance electric fields through mid latitudes to the strong ion outflow regions located at cusp, auroral, and polar cap latitudes. Due to the widespread nature of this system, an inherently global perspective is necessary to describe adequately the key elements in the overall physical picture.

Foster et al. [2005a] discussed a localized enhancement of total electron content (TEC) seen at the base of the erosion plume. Ground based observations with GPS TEC and incoherent scatter radar, and in situ DMSP measurements in the topside ionosphere, suggested that this enhanced TEC results from a poleward redistribution of post-noon sector low latitude ionospheric plasma during the early stages of a strong geomagnetic disturbance. Down-looking IMAGE FUV ionospheric observations [Immel et al., 2005] showed that the TEC enhancement seen from the ground was associated with a localized enhancement somewhat poleward of the equatorial anomaly peak. The enhanced features, seen both from the ground and from space, corotate (approximately) with the Earth once they are formed. These effects are especially pronounced over the Americas, and Foster et al [2005a] suggested that this results from a strengthening of the equatorial ion fountain due to undershielded (penetrating) electric fields in the vicinity of the South Atlantic magnetic anomaly.

Vlasov et al. [2003] considered both observations and model simulations of such a mid- latitude TEC enhancement during the July 15, 2000 storm and concluded that both sunward and poleward convection of the low latitude plasma must be taken into account in addressing such events. They found that the early-event increase of TEC on the dayside could be explained by a small enhancement of the equatorial anomaly combined with the effects of either an equatorward wind or an eastward electric field. However, the large increase in TEC near dusk at mid latitudes indicated both a large enhancement of the equatorial ion fountain (by some combination of dynamo and penetration eastward electric fields) and the effects of a poleward electric field, leading to a sunward and poleward redistribution of plasma. An earlier study by Buonsanto and Foster [1993] examined Arecibo radar observations during a similar event and noted that the high density plasma in the enhanced-TEC region moved approximately horizontally, with a strong poleward and westward component to its motion. For the July, 2000 and similar events, Foster et al. [2005a] described the simultaneous increase in TEC over Florida with a complementary decrease in TEC near the magnetic equator (Brazil) at the edge of the south Atlantic magnetic anomaly (SAA). Further observations and discussion of the redistribution of low-latitude plasma during the July 15, 2000 event have been presented by Yin et al. [2004] and Kelley et al. [2004].

Figure 5-5 presents combined GPS TEC maps of the stormtime perturbations over the Americas and DMSP in-situ observations of topside (~830 km altitude) density during the May 30, 2003 event. A deep electron density/TEC depletion at magnetic equatorial latitudes over Brazil formed near sunset, as in the July, 2000 event described above. Simultaneously, TEC increased dramatically over the northern Caribbean, near the coast of Florida. The SED erosion plume formed at the poleward extent of the mid-latitude TEC enhancement. This progression of events, with the same local time and UT characteristics and spatial localization occurred for a number of the large storm events in the most recent solar cycle maximum. The study of Foster and Coster [2006] found that near sunset during the expansion phase of large geomagnetic storms, a localized enhancement of total electron content (TEC) takes place in

American longitude sector off the coast of Florida and in its magnetically conjugate region. These enhancements lie poleward of the crests of the equatorial ionization anomaly and on field lines mapping into the plasmasphere boundary layer. They concluded that horizontal plasma transport associated with disturbance electric fields is the dominant mechanism for producing the conjugate TEC enhancements. They found these effects to be repeatable in space and time and to create a conjugate intensification of TEC in the outer plasmasphere which serves as a corotating source for the erosion plumes of storm enhanced density which are observed during such events. Figure 5-6, taken from that study, depicts the synchronous decrease in equatorial TEC near the sunset terminator, and simultaneous enhancement of TEC near the PBL at magnetic conjugate points.
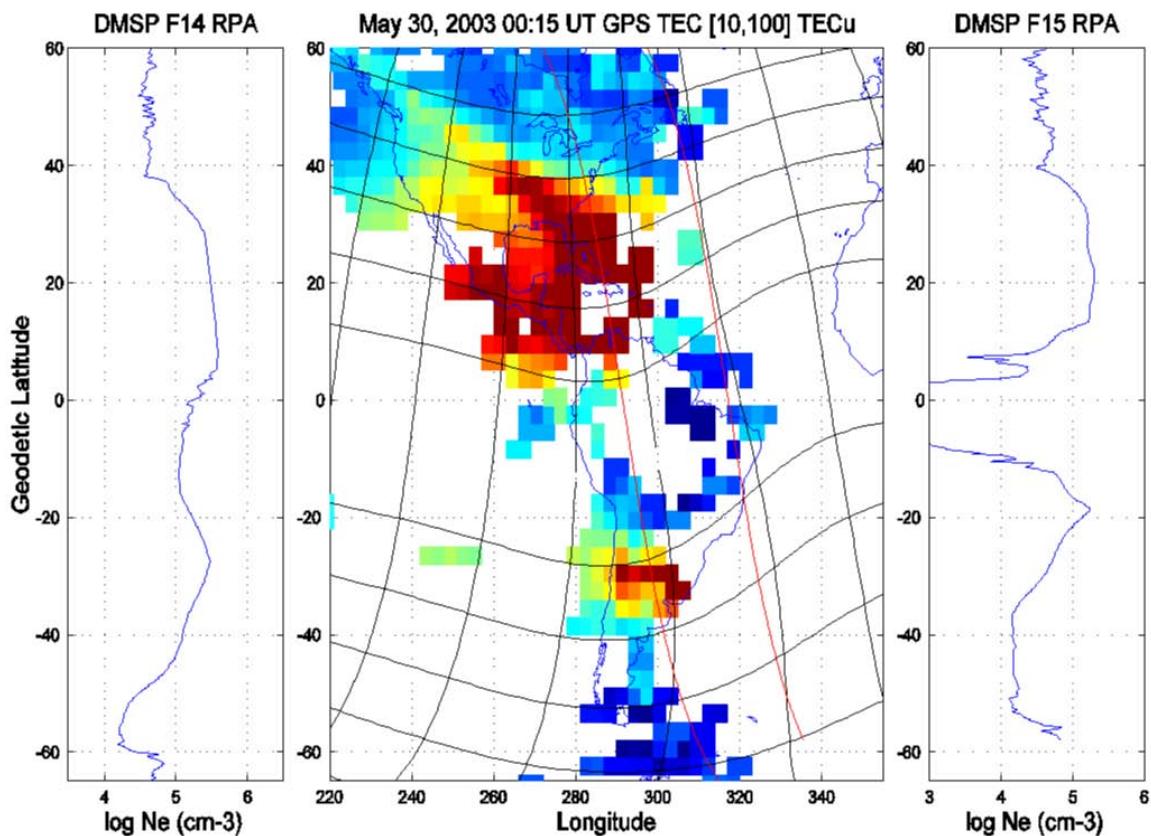
Figure 5-5: Simultaneous GPS TEC and DMSP In Situ Plasma Density Observations Illustrate the Effects of Plasma-Redistribution Shown Schematically in Figure 5-2. The orbital tracks for DMSP F14 (left) and F15 (right) are shown in red. A deep total content hole is formed at the magnetic equator over eastern Brazil while greatly enhanced plasma is relocated to the vicinity of the Florida coast.
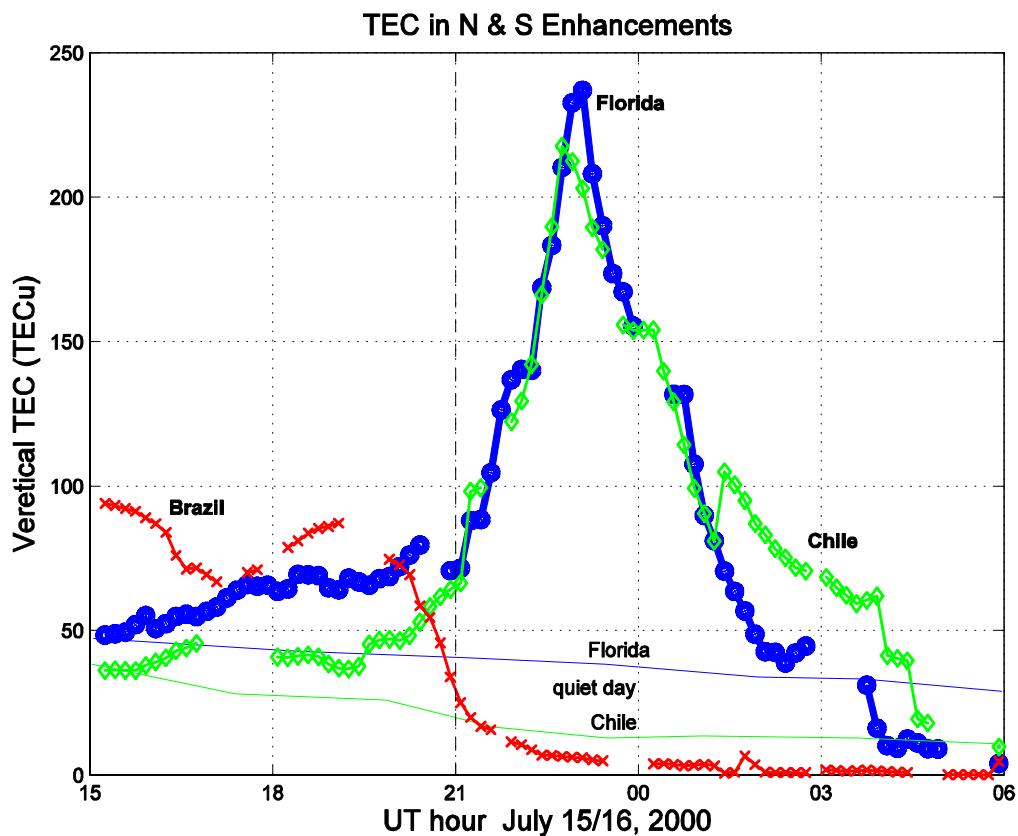
Figure 5-6: The Temporal Variation of TEC in the N-S Conjugate Enhancements During the July 2000 Event Indicates a Close Similarity. TEC over Florida and Chile increase sharply as the equatorial TEC over Brazil drops during storm intensification. Conjugate enhancements of ~ 200 TECu above the quiet-day background (July 13/14) were observed.

5.1.7   Polar Tongue of Ionization

The global character of large-scale ionospheric disturbances during the main phase of a major geomagnetic storm is emphasized by the findings of Foster et al. [2005b]. For the large geomagnetic storm on November 20, 2003, they examined data from the high-latitude incoherent scatter radars at Millstone Hill, Sondrestrom, and EISCAT Tromso, with SuperDARN HF radar observations of the high-latitude convection pattern and DMSP observations of in situ plasma parameters in the topside ionosphere. These were combined with north-hemisphere polar maps of stormtime plumes of enhanced total electron content (TEC) derived from a network of GPS receivers. Figure 5-7, taken from that study, shows the polar tongue of ionization (TOI) to be a continuous stream of dense cold plasma entrained in the global convection pattern. The dayside source of the TOI is the plume of storm enhanced density (SED) transported from low latitudes in the post-noon sector by the sub-auroral disturbance electric field. Convection carries this material through the dayside cusp and across the polar cap to the nightside where the auroral F region is significantly enhanced by the SED material.
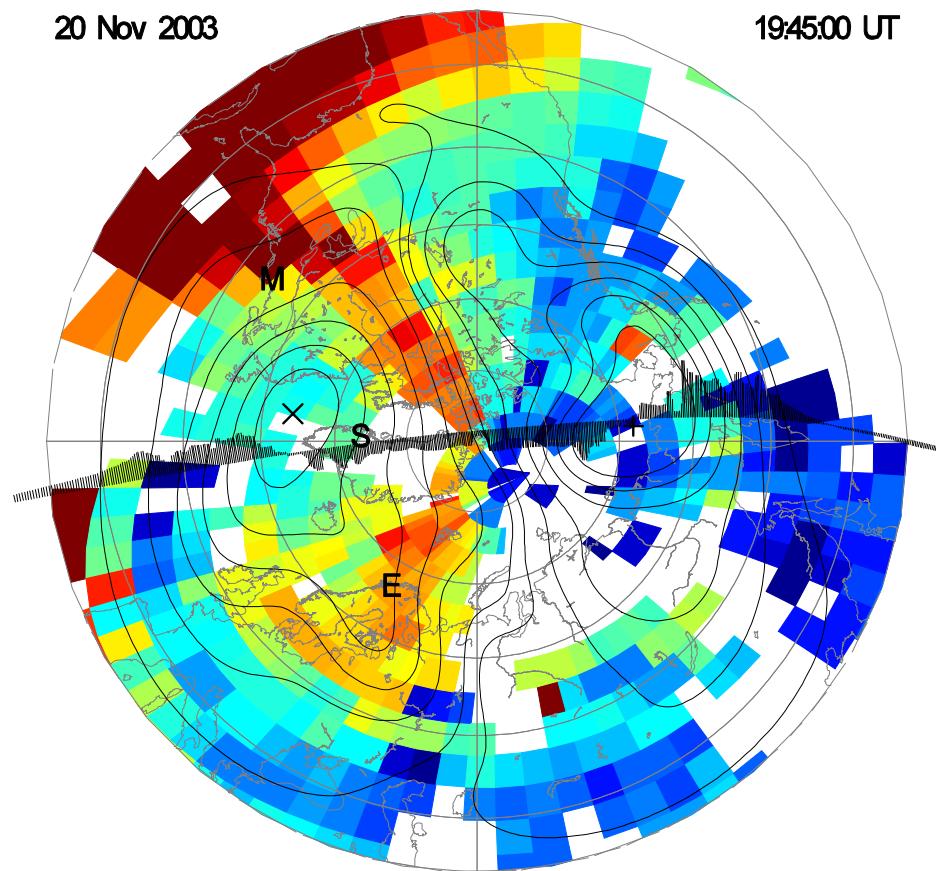
Figure 5-7: Combined GPS TEC and Convection Observations are Displayed in Polar Projection (mag lat/MLT coordinates; 10-deg latitude circles; with noon at the top).

The positions of the three IS radars are indicated (M, S, E). Data are shown for 19:45 UT when the SED/ TOI plume is seen to extend continuously from its low-latitude source in the pre-noon sector, through the dayside cusp and across the polar cap to the midnight sector over the EISCAT facility (E). Vertical TEC observations binned by lat/long at 350 km altitude are displayed with the simultaneous, independent convection pattern derived from combined SuperDARN and DMSP observations. Ion drift meter cross-track velocity data from a trans-polar cap DMSP pass are shown, indicating anti- sunward convection above 60 deg latitude spanning the polar region.

5.1.8   Gradients

Such severe ionospheric perturbations lead to significant space weather effects as steep density gradients and associated regions of ionospheric irregularities form along their edges. Vo and Foster [2001] performed a statistical analysis of 15 years of Millstone Hill ISR observations for the occurrence of density-gradient regions. Density and TEC gradients were observed at sub-auroral and mid latitudes associated with the ionospheric trough and with advecting density enhancements during geomagnetically disturbed conditions. Latitude gradients in TEC associated with the equatorward wall of the trough in the dusk sector were approximately 10 TEC units/degree of latitude for solar max conditions, while gradients associated with regions of greatly-enhanced TEC (storm enhanced density, SED) were observed as great as ~50 TEC units/deg.

Foster and Rideout [2005] used observations from the array of North American GPS receivers to examine the formation and severity of mid-latitude enhancements and steep gradients in total electron content

(TEC) during the October 30-31, 2003 superstorm. A large (~10x) enhancement in dayside TEC was observed over the US mainland during these events as Dst decreased sharply and strong SAPS electric fields eroded the outer reaches of the post-noon plasmasphere boundary layer (PBL) forming poleward-streaming plumes of storm enhanced density. Figure 5-8 shows TEC increased to >250 TECu over the SW USA during the event. TEC gradients across the PBL over the central US (cf. Figure 5-9) exceeded 60 TECu per deg latitude. Basu et al. [2005] identified simultaneous occurrence of scintillations at mid-latitude and in the equatorial region during that event. Mitchell et al. [2005] describe the occurrence of polar-latitude scintillations associated high- gradient regions as the enhanced-TEC material streamed back over the northern polar cap during the October, 2003 event. While intense, the features observed during that superstorm are qualitatively similar to those seen in lesser events. Communications and navigation systems relying on trans-ionospheric propagation must be able to compensate for the effects of such sharp gradients in total electron content in the North American sector.
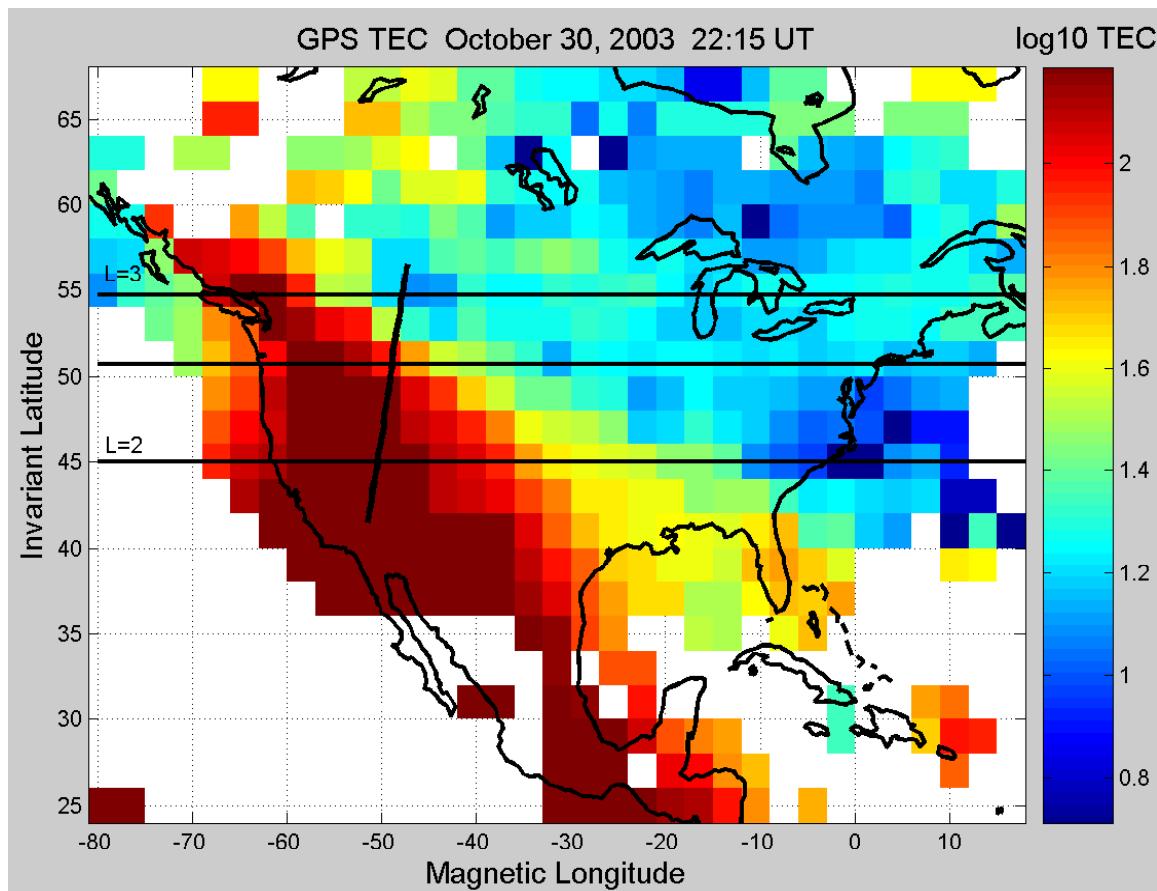
Figure 5-8: TEC Derived from an Array of ~450 GPS Receivers is Displayed in Magnetic Coordinates as a Major Ionospheric Disturbance Forms over the US Mainland During the October 30, 2003 Superstorm. The enhancements of TEC over the US mainland at 22:15 UT is displayed. The SED plume extends into the Pacific NW, TEC exceeds 250 TECu over the US southwest equatorward of L = 2 and the steep gradient region associated with the PBL (cf Figure 5-9).

TEC Gradient at 22:15 UT Oct 30, 2003
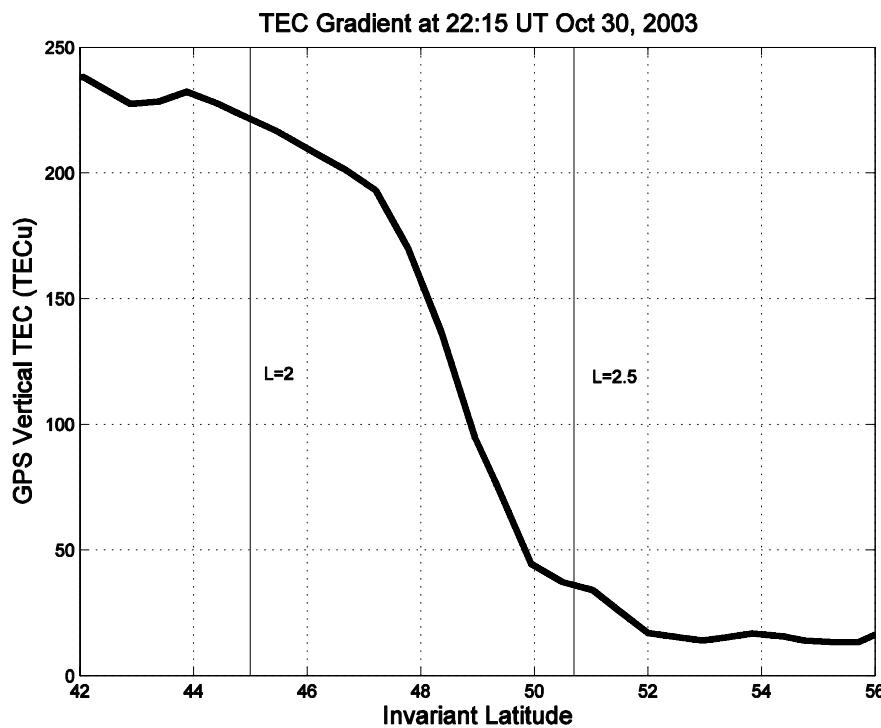
Figure 5-9: The TEC Gradient Across the Poleward Edge of the SED Plume Shown in Figure 5-8 is Presented. TEC in excess of 230 TECu is observed equatorward of the PBL gradients. The steep decrease in TEC with latitude across the PBL extends over ~ 3 deg of latitude with a characteristic gradient of -60 TECu per deg of latitude.

5.1.9 References

Basu, S., Su. Basu, K. M. Groves, E. MacKenzie, M. J. Keskinen, and F. J. Rich, Near-simultaneous plasma structuring in the midlatitude and equatorial ionosphere during magnetic superstorms, *Geophys. Res. Lett.*, 32, L12S05, doi:10.1029/2004GL021678, 2005.

Buonsanto, M. J., A case study of the ionospheric storm dusk effect, *J. Geophys. Res., 100*, 23,857, 1995.

Buonsanto, M. J., and J. C. Foster, Effects of magnetospheric electric fields and neutral winds on the low-middle latitude ionosphere during the March 20-21, 1990 storm, *J. Geophys. Res.*, *98*, 19133-19140, 1993.

Buonsanto, M. J., S. A. Gonzalez, G. Lu, B. W. Reinisch, and J. P. Thayer, Coordinated incoherent scatter radar study of the January 1997 storm, *J. Geophys. Res., 104*, 24,625, 1999.

Carpenter, D. L., and J. Lemaire, 2004. The plasmasphere boundary layer, Annales Geophysicae, 22, 4291.

Chi, P. J., C. T. Russell, J. C. Foster, M. B. Moldwin, M. J. Engebretson, and I. R. Mann, 2005, Density enhancement in the plasmasphere-ionosphere plasma during the 2003 Halloween magnetic storm: Observations along the 265th meridian in North America, *Geophys. Res. Lett.*, 32, L03S07, doi:10.1029/ 2004GL021722.

Evans, J. V., The causes of storm-time increases of the *F*-layer at mid-latitudes, *J. Atmos. Terr. Phys., 35*, 593, 1973.

Foster, J. C., 1993. Storm-Time Plasma Transport at Middle and High Latitudes, Journal of Geophysical Research, 98, 1675-1689.

Foster, J. C., and W. J. Burke, 2002. SAPS: A new characterization for sub-auroral electric fields, EOS, 83, 393-394.

Foster, J. C., and A. J. Coster, Conjugate localized stormtime enhancement of total electron content at low latitudes in the American sector, *J. Atmos. Solar Terr. Phys.*, 69(10-11), 1241-1252, 2007.

Foster, J. C., and F. J. Rich, Prompt midlatitude electric field effects during severe geomagnetic storms, *J. Geophys. Res.*, *103,* 26367, 1998.

Foster, J. C., and W. Rideout, Midlatitude TEC enhancements during the October 2003 superstorm, *Geophys. Res. Lett.*, 32, L12S04, doi:10.1029/2004GL021719, 2005.

Foster, J. C., and H. B. Vo, 2002. Average Characteristics and Activity Dependence of the Subauroral Polarization Stream, *J. Geophys. Res.*, 107(A12),1475, doi: 10.1029/ 2002JA009409.

Foster, J. C., A. J. Coster, P. J. Erickson, J. Goldstein, and F. J. Rich, 2002. Ionospheric Signatures of Plasmaspheric Tails, Geophysical Research Letters, 29(13), 10.1029/2002GL015067.

Foster, J. C., A. J. Coster, P. J. Erickson, F. J. Rich, and B. R. Sandel, 2004. Stormtime observations of the flux of plasmaspheric ions to the dayside cusp/magnetopause, Geophysical Research Letters, 31, L08809, doi:10.1029/2004GL020082.

Foster, J. C., A. J. Coster, P. J. Erickson, W. Rideout, F. J. Rich, T. J. Immel, and B. R. Sandel, 2005a, Redistribution of the stormtime ionosphere and the formation of the plasmaspheric bulge, in *Inner Magnetosphere Interactions: New Perspectives from Imaging*, 2004 Yosemite Conf. Proceedings, AGU Press, Wash. DC.

Foster, J. C., A. J. Coster, P. J. Erickson, J. M. Holt, F. D. Lind, W. Rideout, M. McCready, A. van Eyken, R. A. Greenwald, F. J. Rich, 2005b. Multiradar observations of the polar tongue of ionization, *J. Geophys. Res.*, 110, A09S31, doi:10.1029/2004JA010928.

Foster, J. C., W. Rideout, B. Sandel, W. T. Forrester, and F. J. Rich, 2006, On the relationship of SAPS to storm enhanced density, J. Atmos. Space Terr. Phys., in press.

Gonzales, C. A., M. C. Kelley, L. A. Carpenter, and R. H. Holzworth, Evidence for a magnetospheric effect on mid-latitude electric fields, *J. Geophys. Res., 83*, 4397, 1978.

Huang, C. S., J. C. Foster, and M. C. Kelley, Long-duration penetration of the interplanetary electric field to the low-latitude ionosphere during the main phase of magnetic storms, *J. Geophys. Res.*, *110*, A11309, doi:10.1029/2005JA011202, 2005.

Huang, C. S., J. C. Foster, L. P. Goncharenko, P. J. Erickson, W. Rideout, and A. J. Coster, A strong positive phase of ionospheric storms observed by the Millstone Hill incoherent scatter radar and global GPS network, *J. Geophys. Res.*, *110,* A06303, doi:10.1029/2004JA010865, 2005.

Huang, C. S., S. Sazykin, J. L. Chau, N. Maruyama, and M. C. Kelley, Penetration electric fields: Efficiency and characteristic time scale, *J. Atmos. Sol. Terr. Phys., 68*, in press, 2006.

Immel T. J., J. C. Foster, A. J. Coster, S. B. Mende, H. U. Frey (2005), Global storm time plasma redistribution imaged from the ground and space, Geophys. Res. Lett., 32, L03107, doi:10.1029/ 2004GL021120, 2005.

Jaggi, R. K., and R. A. Wolf, Self-consistent calculation of the motion of a sheet of ions in the magnetosphere, *J. Geophys. Res.*, *78*, 2852, 1973.

Kelley, M. C., B. G. Fejer, and C. A. Gonzales, An explanation for anomalous equatorial ionospheric electric field associated with a northward turning of the interplanetary magnetic field, *Geophys. Res. Lett., 6*, 301, 1979.

Kelley, M. C., M. Vlassov, J. C. Foster, and A. J. Coster, 2004, A quantitative explanation for the phenomenon known as plasmaspheric tails or storm-enhanced density, *Geophys. Res. Lett.*, 31, L19809, doi:10.1029/2004GL020875.

Mannucci, A. J., B. T. Tsurutani, B. A. Iijima, A. Komjathy, A. Saito, W. D. Gonzales, F. L. Guarnieri, J. U. Kozyra, and R. Skoug, Dayside global ionospheric response to the major interplanetary events of October 29-30, 2993 "Halloween Storm", *Geophys. Res. Lett.*, 32, L12S02, doi:10.1029/2004GL021467, 2005.

Mendillo, M., A study of the relationship between geomagnetic storms and ionospheric disturbance at mid-latitudes, *Planet. Space Sci., 21*, 349, 1973.

Mitchell, C. N., L. Alfonsi, G. De Franceschi, M. Lester, V. Romano, and A. W. Wernik, GPS TEC and scintillation measurements from the polar ionosphere during the October 2003 storm, *Geophys. Res. Lett.*, 32, L12S03, doi:10.1029/2004GL021644, 2005.

Rowland, D. E., and J. R. Wygant, Dependence of the large scale, inner magnetospheric electric field on geomagnetic activity, *J. Geophys. Res.*, 103, 14959, 1998.

Sandel, B. R., R. A. King, W. T. Forrester, D. L. Gallagher, A. L. Broadfoot, and C. C. Curtis, 2001. Initial Results from the IMAGE Extreme Ultraviolet Imager, Geophysical Research Letters, 28, 1439.

Tsurutani, B. T., Mannucci A., Iijima B., Abdu M. A., Sobral J. H. A., Gonzalez W., Guarnieri F., Tsuda T., Saito A., Yumoto K., Fejer B., Fuller-Rowell T. J., Kozyra J., Foster J. C., Coster A., Vasyliunas V. M., 2004, Global dayside ionospheric uplift and enhancement associated with interplanetary electric fields, J. Geophys. Res., 109, A08302, doi:10.1029/2003JA010342.

Vo, H. B, and J. C. Foster, A Quantitative Study of Ionospheric Density Gradients at Mid- Latitudes, *J. Geophys. Res. 106*, 21555-1563, 2001.

Yeh, H.-C., J. C. Foster, F. J. Rich, and W. Swider, 1991. Storm-time electric field penetration observed at mid-latitude, Journal of Geophysical Research, 96, 5707.

Yin P., C. N. Mitchell, P. S. J. Spencer, J. C. Foster, 2004, Ionospheric electron concentration imaging using GPS over the USA during the storm of July 2000, *Geophys. Res. Lett.*, 31, L12806, doi:10.1029/ 2004GL019899.

5.2 IONOSPHERIC PROCESSES OF IMPORTANCE TO SEVERE SPACE WEATHER: A REVIEW[1]

(Mike Kelley, Cornell University)

5.2.1 Introduction

Fifty years into the Space Age, technical societies now are deeply committed to utilizing space. For the military, space is the ultimate high ground from which a variety of surveillance, communications, and navigation systems operate. For industry, the communications and positional/navigational opportunities

---

[1] Adapted from *Kelley et al.* [2006]. Reproduced with permission of the American Geophysical Union.

for using space-based systems are virtually unlimited. However, when plasma between the satellite and the receiver is turbulent, satellite signals scintillate, similar to the twinkling of starlight as it traverses the turbulent atmosphere, and communications and navigation systems can be seriously affected. Additionally, when steep gradients in the plasma content occur, navigation solutions are very difficult to achieve.

The natural assumption is that these problems maximize in the auroral zone where the energy input from the solar wind is at its peak. However, the most severe ionospheric weather occurs within ±20º of the geomagnetic equator where stored gravitational energy sometimes is released after sunset [1]. Depending on the condition of the equatorial ionosphere at that time, a null or modest event may occur or, more problematically, vast convective plumes of turbulent plasma can erupt, devastating communications from HF to GHz frequencies [2,3]. The plumes rise rapidly in a manner analogous to thunderstorm convection and, due to their electrical properties, are transmitted rapidly for vast distances north and south along the earth's magnetic field. Termed Convective Ionospheric Storms (CIS) to emphasize the analogy of the ionospheric process to thunderstorms, these plumes are caught up in the high-speed eastward plasma drift and often last until well after midnight, meaning that a single structure can affect a very large area of the earth in its lifetime [4].

Another surprising and more recent result is that very sharp horizontal gradients can develop at middle latitudes at night and sometimes even in the daytime [5]. Most of the ones observed to date developed over the Caribbean and erupted over CONUS, but they have also been seen in the European sector. These gradients can seriously affect systems such as the Wide Area Augmentation System (WAAS), which is based on GPS [6].

The physics behind these two phenomena are discussed next.

5.2.2   Convective Ionospheric Storms

The morphology of the equatorial ionosphere is quite different from that at other latitudes because the magnetic field is nearly parallel to the earth's surface. During the daytime, the plasma drifts upward under the influence of an eastward-directed electric field. The uplifted plasma then moves along the magnetic field in response to gravity and pressure-gradient forces, producing a fountain effect in which plasma moves upward close to the equator and then slides down the field line on the north and south sides. Two ionization maxima are formed on either side of the equator – the so-called equatorial Appleton anomalies, which create particularly dense plasma, ripe for disturbing satellite signals. The two crests are approximately ±15º from the magnetic equator [2].

The equatorial electric field is westward (eastward) during the night (day) and the plasma drifts downward (upward) in a diurnal pattern. However, just before this electric field reversal takes place, the upward drift increases suddenly, driving the main dense ionosphere, called the F layer, to high altitudes. In addition, production of ions by sunlight ceases after sunset and the conductivity of the lower ionosphere decreases dramatically, reducing its capability for "shorting out" the electrical potential. This large upward drift, combined with gravity, a decreasing neutral atmospheric density, and a vertical plasma-density gradient in the bottomside of the F layer, creates an unstable configuration. If the ionosphere is low in altitude on a given night, the event usually fades out. But if the post-sunset, upward-surging plasma layer is high enough (at or above 350 km), the stored energy can be released explosively, forming the low density, highly turbulent features that surge upward [7].

Like a thunderstorm, CIS releases gravitational energy stored in the height of the nighttime ionosphere, making altitude an important factor in the strength of the CIS – the higher the post-sunset ionosphere, the more gravitational potential it accumulates and the more likely the storm will be severe. The events usually die out by midnight but occasionally regenerate later, an occurrence that is associated with high geomagnetic activity [2].

During a severe event, the electric fields generated in the low density regions drive plumes of turbulent, low density plasma upward to heights of over 2000 km at the equator. Due to the high conductivity of plasma parallel to the magnetic field, these equatorial structures connect electrically to the plasma over good viewing sites such as Hawaii and Cerro Tololo, Chile to form mirroring features by CCD cameras [8]. In this manner, airglow emissions are used to trace the development of the event over vast distances from these sites and similar locations around the world. These features are now detectable from on-orbit systems such as the one flown on the TIMED satellite, a development that eventually could lead to a space weather observatory at geosynchronous orbit [9].

In the near future, a joint project between the Air Force Research Laboratory and the DOD Space Test Program, called the Communications/Navigation Outage Forecasting System (C/NOFS), has the mission goal of predicting CIS and soon will provide detailed, three-dimensional images of these storms [3]. To accomplish this task, the project will gather the most complete set of in situ and ground-based data ever attempted. These data sets include on-orbit and/or ground-based measurements of electric and magnetic fields, neutral winds, photon emissions, Doppler radar images, GPS and lower frequency observations of scintillations, GPS occultation observations, tomographic maps of the ionosphere, plasma density and its fluctuations, and total electron content. These data will be assimilated into predictive physics-based models with the goal of predicting scintillation for Air Force systems [3].

However, although space weather scientists have a solid grasp of the underlying physics of CIS, understanding the day-to-day variability of this phenomenon has proven difficult. The major challenge is to understand the physical processes leading to the formation of plasma irregularities in the ionosphere and to identify the mechanisms triggering or inhibiting plasma instability. Meeting this challenge requires accurate modeling of the parameters that are part of the instability growth rate. The electric field, whether caused by neutral wind (the dynamo electric field) or by magnetospheric phenomena, is probably the most important of these parameters [7]. Since the height of the ionosphere immediately after sunset is a major factor in determining whether or not CIS will occur, knowing the eastward electric field component driving the upward plasma motion is vital to our predictive capability. The electric field measurements obtained by the C/NOFS satellite will provide crucial information on this key parameter.

Other, more vexing issues involve lower atmospheric tides and waves that create initial disturbances [10] and the electrical conductivity of the low latitude ionosphere in contact with the rising ionosphere at sunset [11], which may or may not short-circuit the electric fields. The collisional Kelvin-Helmholtz instability also can seed CIS [12]. The physics-based modeling of plasma bubbles – how they are born, how they evolve in time and space, and how they die – also needs considerable improvement.

5.2.3   Steep Mid-Latitude Gradients

The problems associated with mid-latitude gradients are also severe and were even more unexpected than those associated with CIS [6]. One source is the development of mesoscale Travelling Ionospheric Disturbances (TIDs), which draw energy from either neutral winds or solar wind-related penetrating electric fields [13,14]. The initial TIDs then deepen and, concurrently, one edge steepens to form sharp gradients while the other edge becomes turbulent. For example, one such event was associated with a 15 Total Electron Content (TEC) unit difference as measured by two separated receivers looking at the same puncture point in the ionosphere [15].

Other substantial mid-latitude events seem to be related to intense magnetic storms and are referred to as Storm-Enhanced Densities (SED) [5]. The origin of the very dense plasma seems to be the equatorial daytime ionosphere, which can create copious amounts of plasma under full sunlight that is then continuously pumped up into the equatorial fountain and expands over the middle latitudes. These regions also steepen and exhibit the highest TEC values ever recorded, about 200 TEC units. The storm seems to generate a variety of electrical changes related to the SED events [16]. First, penetrating electric fields of

solar wind origin reach the magnetic equator with an eastward zonal component as the region 1 currents (highest latitude field-aligned currents) intensify, causing the storm time fountain effect mentioned above [17]. Then penetrating meridional electric fields drive the anomaly plasma into the daytime region where it is caught up in the classic convection pattern, established by then deep into the plasmasphere. The plume of high TEC plasma eventually flows anti-sunward into the polar cap [5].

5.2.4   Other Topical Research Areas

Global Thermospheric Winds During Storms. Studies of neutral atmospheric changes due to high latitude heating and momentum transfer have not been as successful as expected over the years. Current thinking is that small scale structure in the electric field contributes as much Joule heating as the average field because the fluctuating component is comparable to the average [2]. With the recent deployment of the Advanced Modular Incoherent Scatter Radar (AMISR) in Alaska and the rocket projects aimed at understanding this problem, we should soon have a better understanding of this important coupling issue. The thermospheric circulation changes so drastically that the worldwide electric field distribution is also affected. This is called the disturbance dynamo and penetrates all the way to the magnetic equator [18]. Once there, this phenomenon can also affect CIS development.

Development of Physics-Based Assimilative Models. Today, space weather scientists are building assimilative physics-based models like those used in weather forecasting, but currently the models are starved for data. The meteorological community uses hundreds of balloon soundings, made twice daily around the world, in their assimilative models. The ever-growing numbers of GPS receiving stations on the ground and on orbiting spacecraft are beginning to provide ionospheric data similar in density to the balloon soundings in meteorology. For example, the two GPS signals transmitted on every satellite can be combined to measure the total number of electrons between the satellite and the receiver. TEC information will be one of the key data streams assimilated in the models now being developed for predicting severe space weather [19]. In addition, with the help of a U.S. firm, Taiwan has just developed and launched six satellites in a system called COSMIC that is capable of over 2500 observations of GPS satellite occultations each day and potentially will provide plasma density profiles as inputs for data assimilation models for Space Weather, models that are analogous to meteorological models [20].

Active Experiments in the Ionosphere. High power radiowave systems, such as the one at the heart of the High-frequency Active Auroral Research Program (HAARP), continue to be of great interest. For the first time, E-region emissions have been generated at both low and high latitudes [20,21]. The latter was so intense that it reached levels visible to the naked eye, although at this point still too small (<1 km) to be seen [23]. Bright spots have also been produced in the F region when the beam is nearly parallel to the magnetic field. Research into generating ELF and VLF signals continues. The latter is of interest for possible radiation belt control in an emergency. Space Shuttle burns, which create significant depletions in the ionosphere, are planned over ground sites in the near future [24]. A near kiloton explosion is also being considered to determine detectability using ionospheric sensors (W. Junior, personal communication, 2006).

5.2.5   Conclusions

Ionospheric research continues to be scientifically rich and has increasing relevance in using space for commercial and military systems. The exploratory era is nearing its end, although certainly new results are reported frequently, and a new emphasis on prediction is underway. The Communications/Navigation Outage Forecast System is a prime example and, with the associated satellite launch in 2008, will play a major role in testing these waters. Building partnerships with countries that are developing strong space science programs is crucial. Japan has played a major role in Solar X-ray studies; rocket, airglow, and radar studies at mid-latitudes; and, of course, the International Space Station. Taiwan has also been very active with the ROCSAT 1 ionospheric sensor package and the recent COSMIC constellation.

One crucial area that may require similar cooperation is solar wind monitoring. If NASA does not continue this crucial role, much of our progress in Space Weather may grind to a halt.

5.2.6   Acknowledgments

5.2.7   References

[1] Kelley, M.C., J.J. Makela, and O. de La Beaujardiere (2006), Convective ionospheric storms: A major space weather problem, *Space Weather, 4*(2), S02C04, doi:10.1029/ 2005SW000144.

[2] Kelley, M.C. (1989), *The Earth's Ionosphere: Electrodynamics and Plasma Physics*. Academic Press, Inc., San Diego, Calif.

[3] de la Beaujardiére, O., J. M. Retterer, M. C. Kelley, D. Hunton, and L. Jeong (2006), New satellite will forecast ionospheric disturbances, *Space Weather, 4*, S03C05, 10.1029/ 2005SW000145.

[4] Basu, Su., Sa. Basu, J. Aarons, J. P. McClure, and M. D. Cousins (1978), On the coexistence of kilometer- and meter-scale irregularities in the nighttime equatorial *F*-region, *J. Geophys. Res., 83,* 4219.

[5] Foster, J. C., P. J. Erickson, A. J. Coster, J. Goldstein, and F. J. Rich (2002), Ionospheric signatures of plasmaspheric tails, *Geophys. Res. Lett., 29*(13), 1623, doi:10.1029/ 2002GL015067.

[6] Coster, A. J., E. M. Gaposchkin, and L. E. Thornton (1992), Real-time ionospheric monitoring system using GPS, *Navigation, 39*(2), 191-204.

[7] Fejer, B. G., L. Scherliess, and E. R. de Paula (1999), Effects of the vertical plasma drift velocity on the generation and evolution of equatorial spread *F, J. Geophys. Res., 104,* 19,859.

[8] Kelley, M. C., J. J. Makela, B. M. Ledvina, and P. M. Kintner (2002), Observations of equatorial spread-F from Haleakala, Hawaii, *Geophys. Res. Lett.*, *29*(20), doi:10.1029/ 2002GL015509.

[9] Kelley, M. C., J. J. Makela, L. J. Paxton, F. Kamalabadi, J. M. Comberiate, and H. Kil (2003), The first coordinated ground- and space-based optical observations of equatorial plasma bubbles, *Geophys. Res. Lett., 30*(14), 1766, doi:10.1029/2003GL017301.

[10] Kelley, M. C., M. F. Larsen, C. A. LaHoz, and J. P. McClure, (1981), Gravity wave initiation of equatorial spread *F*: A case study, *J. Geophys. Res., 86,* 9087.

[11] Kelley, M.C., V. K. Wong, G. A. Hajj, and A. J. Mannucci (2004), On measuring the off-equatorial conductivity before and during convective ionospheric storms, *Geophys. Res. Lett., 31,* L17805, doi:10.1029/2004GL020423.

[12] Hysell, D. L., and E. Kudeki, (2004), Collisional shear instability in the equatorial F region ionosphere, *J. Geophys. Res., 109*, A11301, doi:10.1029/2004JA010636.

[13] Fukao, S., M. C. Kelley, T. Shirakawa, T. Takami, M. Yamamoto, T. Tsuda, and S. Kato, (1991), Turbulence upwelling of the mid-latitude ionosphere, 1. Observational results by the MU radar, *J. Geophys. Res., 96*, 3725-3746.

[14] Saito, A., T. Iyemori, L. G. Blomberg, M. Yamamoto, and M. Takeda (1998), Conjugate observations of the mid-latitude electric field fluctuations with the MU radar and the Freja satellite, *J. Atmos. Solar-Terr. Phys., 60*(1), 129-140.

[15] Makela, J. J., M. C. Kelley, J. J. Sojka, X. Pi, and A. J. Mannucci (2001), GPS normalization and preliminary modeling results of total electron content during a midlatitude space weather event, *Radio Sci., 36*(2), 351-362.

[16] Foster, J. C., and H. B. Vo (2002), Average characteristics and activity dependence of the subauroral polarization stream, *J. Geophys. Res.*, *107*(A12), 1475, doi:10.1029/2002JA009409.

[17] Foster, J. C., P. J. Erickson, F. D. Lind, and W. Rideout (2004), Millstone Hill coherent-scatter radar observations of electric field variability in the sub-auroral polarization stream, *Geophys. Res. Lett.*, *31*, L21803, doi:10.1029/2004GL021271.

[18] Fejer, B. G., M. F. Larsen, and D. T. Farley (1983), Equatorial disturbance dynamo electric fields, *Geophys. Res. Lett., 10*, 537.

[19] Zhu, L., R. W. Schunk, G. Jee, L. Scherliess, J. J. Sojka, and D. C. Thompson (2006), Validation study of the Ionosphere Forecast Model using the TOPEX total electron content measurements, *Radio Sci., 41*, RS5S11, doi:10.1029/2005RS003336.

[20] http://www.nspo.org.tw/2005e/projects/project3/hot.shtml.

[21] Djuth, F. T., P. A. Bernhardt, C. A. Tepley, J. A. Gardner, M. C. Kelley, A. L. Broadfoot, L. M. Kagan, M. P. Sulzer, J. H. Elder, C. Selcher, B. Isham, C. Broadfoot, and H. C. Carlson (1999), Large airglow enhancements produced via wave-plasma interactions in sporadic *E, Geophys. Res. Lett., 26*(11), 1557-1560.

[22] Kagan, L. M., M. C. Kelley, F. Garcia, P. A. Bernhardt, F. T. Djuth, M. P. Sulzer, and C. A. Tepley (2000), The structure of electromagnetic wave-induced 557.7 nm emission associated with a sporadic-*E* event over Arecibo, *Phys. Rev. Lett., 85*(1), 218-221.

[23] Pedersen, T. R., and Gerken, E. A. (2005), Creation of visible artificial optical emissions in the aurora by high-power radio waves, *Nature, 433*, 498-500, doi:10.1038/nature03243.

[24] Bernhardt, P. A., J. D. Huba, E. Kudeki, R. F. Woodman, L. Condori, and F. Villanueva (2001), Lifetime of a depression in the plasma density over Jicamarca produced by space shuttle exhaust in the ionosphere, *Radio Sci., 36*(5), 1209-1220, 10.1029/2000RS002434.

5.3   TECHNOLOGY AND SECURITY AFFECTED BY SPACE WEATHER

(Frank Jansen, 1A – First Applications on Space Weather Service Greifswald & Space Weather Observatory Greifswald, 17498 Greifswald – Groß Karrendorf, Germany)

Abstract

This report describes the background of space weather development in the U.S.A., in Europe (EU and ESA) and highlights of the developments (Introduction of this section). The second section contains an overview and conclusions for each known affected technology branch. The third section describes several security aspects, latest developments and conclusions related to space weather.

5.3.1   Introduction

Several space weather effects related to technology, important for society, have been reported in the last decade ([1] and [2]).

Specific funding of space weather observations started in the U.S.A. in the early 1990s. These activities were partially a result of the superstorm and power blackout on 13 March 1989, as well as several satellite failures and communication problems. The National Space Weather Program (NSWP, under www.nswp.gov) started in 1994 instigated by several U.S. government agencies to guard against space weather effects. The overall goal of the NSWP is to achieve an active, synergistic, interagency system to deliver timely, accurate and reliable space weather warnings, observations, specifications and forecast. The following U.S. organizations contribute to NSWP: NSF, NASA, DoD, DoC, DoE, DoI and DoT. The second edition of the NSWP implementation plan of July 2000 describes in detail the direction for research, operations, education, training and program management to achieve the goals of NSWP until the end of the first decade of the 21st century.

In Europe space weather activities started in the second half of the 1990s. The first workshop on space weather was organized by ESA in 1998 [3]. It had sessions on effects and users, physical processes, models and data. Various worldwide space weather initiatives resulted from deliberations and discussions. There was general support for the idea of a European Space Weather Program (ESWP). Two parallel ESA space weather feasibility studies were carried out between 1999 and 2001 by the Alcatel Space Industry (France) and Rutherford Appleton Laboratories (RAL, UK) consortia.
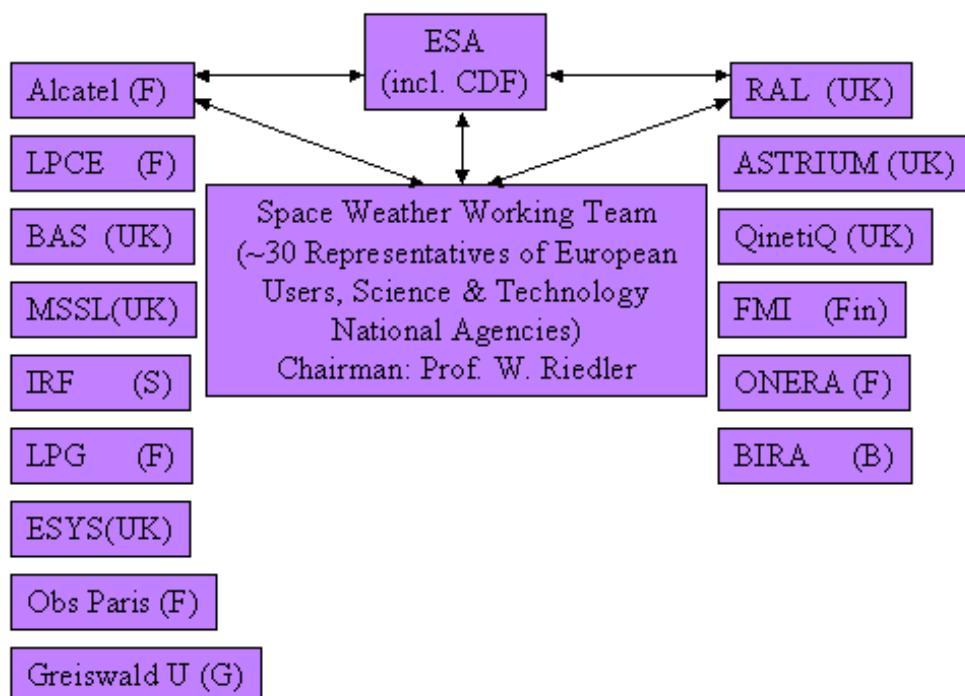

**Figure 5-10: Main Bodies of Early Space Weather Activity in Europe:
ESA, members of the Alcatel and RAL teams, Space Weather Working
Team (SWWT) and ESA's Concurrent Design Facility (CDF).**

The studies included for example:

1) The benefits, detailed rationale and details of an ESWP, including space and ground segment and prototyping of services;

2) The definition of structures that must be implemented by ESA and its member states; and

3) A draft programme, cost and risks.

The available ground-based and space-based instruments in Europe and the necessary capabilities for information about the Sun, interplanetary space, the magnetosphere, the ionosphere and the thermosphere were reviewed. For example, the status of ionospheric tools such as SuperDARN coherent radars, ionosondes, incoherent scatter radars were judged as in research status, but positional receivers were seen to be operational for GPS applications. In addition a cosmic ray muon telescope – required in the study of the Alcatel team – for the European region was also proposed to forecast the onset of disturbances in the magnetosphere/ionosphere. MuSTAnG (Muon Spaceweather Telescope for Anisotropies at Greifswald) was submitted for funding by ESA and DLR [4] and constructed from 2005 to 2006. It is currently in a test phase and will measure in real time the propagation of interplanetary Coronal Mass Ejections (CME) and forecast their arrival time at Earth.

Figure 5-11 shows the first European space weather telescope MuSTAnG (Muon Spaceweather Telescope for Anisotropy at Greifswald) which will assist in forecasting space weather storms.
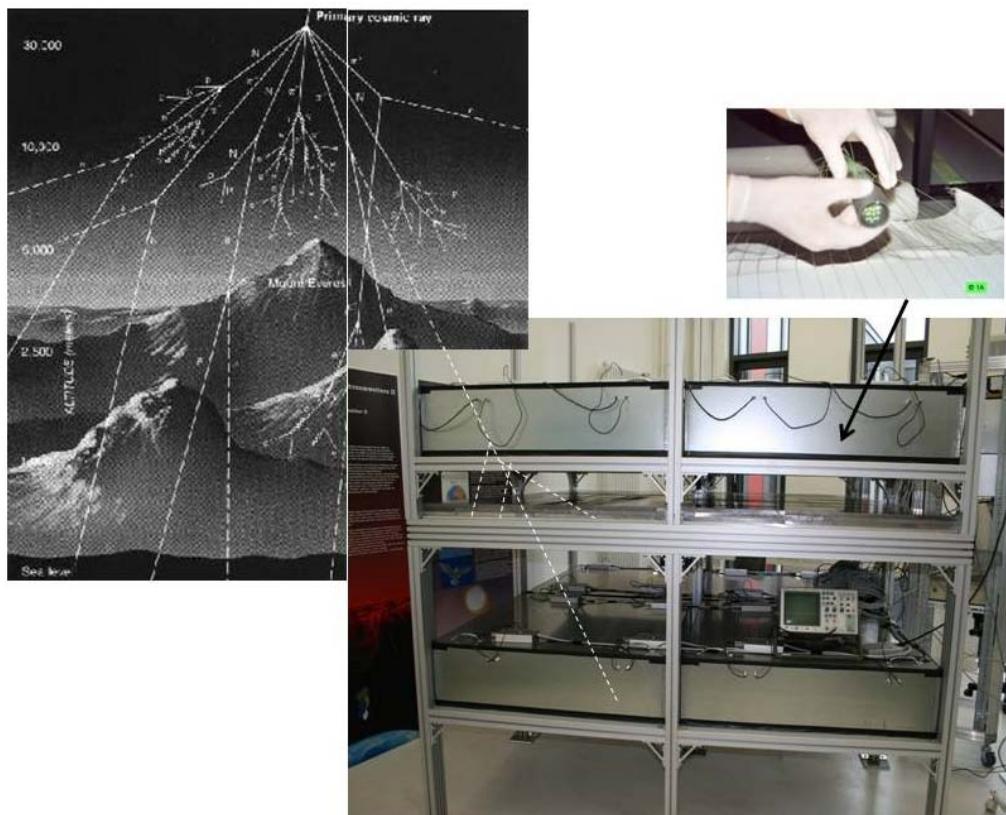
Figure 5-11: The Muon Space Weather Telescope for Anisotropy at Greifswald (MuSTAnG).

White lines signify cosmic rays in the Earth atmosphere and the MuSTanG telescope. Only muons reach the upper and lower detector layer. Other particles are stopped in a lead layer. Muons create flashes of green light in the Mustang telescope scintillator plates. The plates and fibres are mounted inside light-tight metal boxes (image: F. Jansen, 1A).

For the space-based segment an entire fleet of spacecraft and satellites including instrumentation, orbits and communication links are proposed finally by the two study consortia Alcatel and RAL to ESA.

Documents about space and ground segment, space weather effects and parameters, user requirements and market analysis as well as about the rationale and recommendations for the ESWP are available online at: http://esa-spaceweather.net/spweather/esa_initiatives/spweatherstudies/public_doc.html. In addition, the Space Weather Working Team (SWWT) was created, with about 30 European experts in a variety of both scientific and application-oriented fields relating to space weather. The team members also evaluated the primary scientific orientated results of both consortia. The consortia proposed many interesting options for a space weather system (see also Figure 5-12 and Figure 5-13 below). ESA's Concurrent Design Facility (CDF) further analysed these options to establish their feasibility and costs.
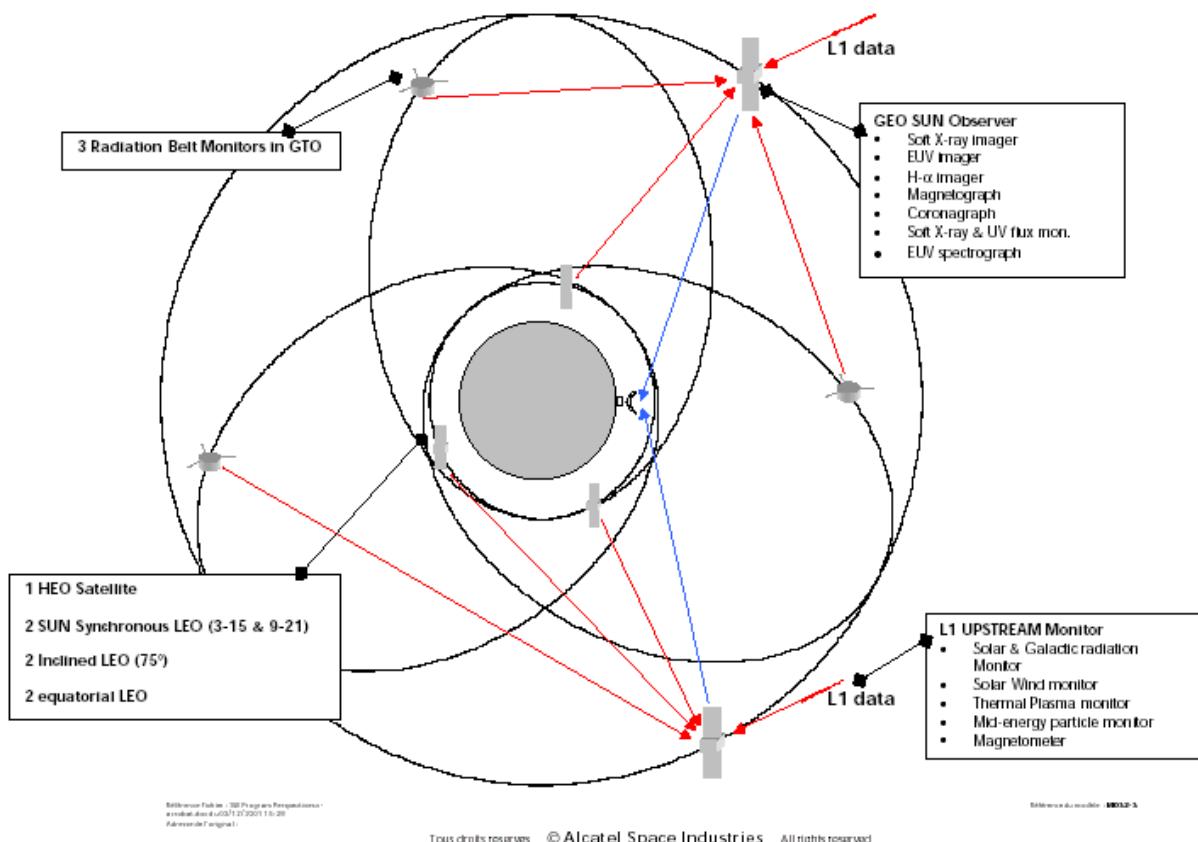

Figure 5-12: Full Space-Based Scenario by the Alcatel Space Industry Consortium with Global Data Communication Consisting of: Sun Observer and Upstream Monitor at L1, Three Radiation Belt Monitors in Geostationary Transfer Orbit (GTO) and Seven Additional Satellites in High Eccentric Orbit (HEO) and Low Earth Orbit (LEO).
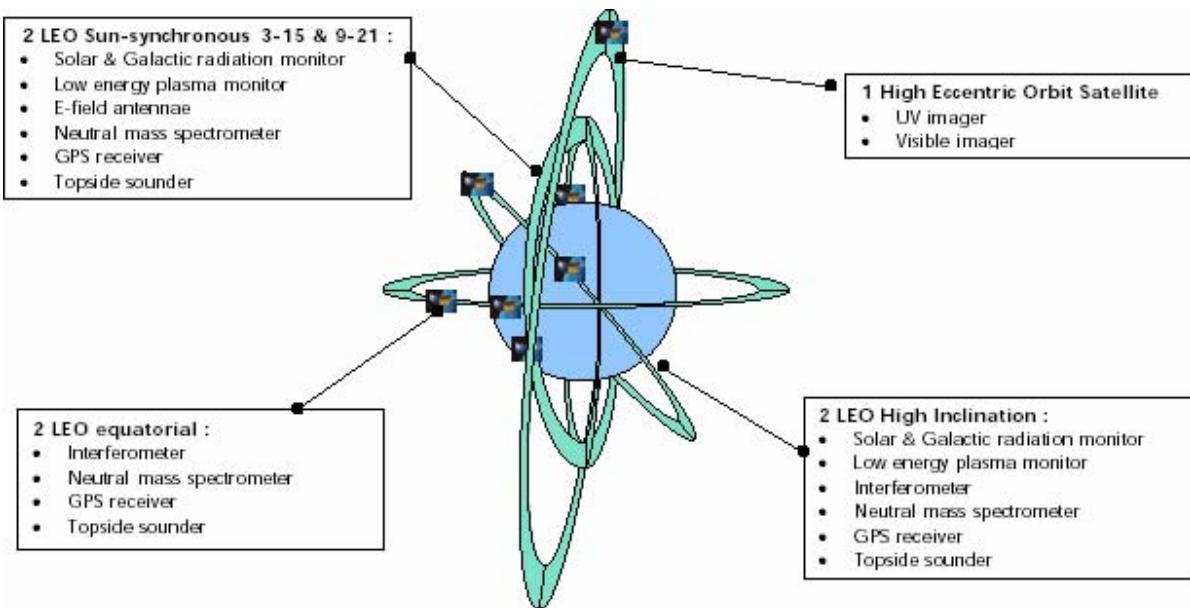
Figure 5-13: The Alcatel Space Industry Consortium Proposed this Space-Based Full Scenario of Satellites and Instruments in HEO and LEO; in Sun-Synchronous, Equatorial and High Inclination Orbits in the Ionosphere and Thermosphere.

On the European level several workshop proceedings, books and DVDs were published related to space weather: "Space Storms & Space Weather Hazards" in Greece [5] in 2000, and the SOLSPA Euroconference "Solar Cycle and Space Weather" in Italy [6] in 2001. After 1998, the year of the first workshop on space weather in Europe, ESA has organised space weather workshops each year. There was a workshop on the topic of the utilization of a future European space weather service [7] in 2000. The 2001 workshop looked towards a future European space weather programme [8]. In 2002 a potential space weather applications pilot project came into discussion [9]. In 2003 the workshop was orientated towards developments of a European space weather service network [10]. In 2002 the European Commission funded European Science Week, which carried out the project SWE (Space Weather and Europe) to promote societal and public awareness aspects of space weather all over Europe: SWE organised a public space weather forum in Berlin including a real-time video link to the International Space Station for about 300 politicians and representatives from industry and the media. The first interactive space weather CD-Rom worldwide was published by the SWE project in English, French and German [11]. A new level of space weather activities started with the European Space Weather Weeks in 2004 [12] and 2005 [13] at ESA/ESTEC in the Netherlands. To promote space weather aspects for the European Commission the space weather weeks in 2006 and 2007 were organised in Brussels. The first European space weather fair was a part of the second public space weather forum in Schwerin during the European Science Week in 2007. It was organised within the EU-funded SWEETS project (Space Weather and Europe – an Educational Tool with the Sun, see under www.sweets2007.eu), a successor of SWE. SWEETS published a new, educational space weather DVD in 8 languages [2].

Moreover, the European Commission and ESA published an action plan for implementing the European Space Policy in 2003 [14]. This White Paper underlined specific efforts to ensure that Europe has the capacity to supply to the different users critical information on space weather predictions.

In 2003 the European Science Foundation started the 'Action 724 on the European Cooperation in the field of Scientific and Technical Research (COST)'. This developed the scientific basis for monitoring, modelling and predicting space weather. The main aims of the action were:

1) To coordinate European research into modelling and prediction of space weather;

2) To promote new instrument deployment to satisfy data requirements and development of new models; and

3) To educate potential users.

The main benefit of this concerted space weather action towards an ESWP is that organisations will have a resource which will enable them to manage space weather risks. Especially the development of a coordinated European space weather network relevant to data, models, prediction and public outreach and education is foreseen. An international advanced school on space weather was initiated by the COST 724 action and realized in 2006 at the Abdus Salam International Centre for Theoretical Physics in Trieste [16]. The COST 724 action objectives described as tasks of the four working groups "Monitoring and predicting solar activity for space weather", "The radiation environment of the Earth", "Interaction of solar wind disturbances with the Earth" and "Space Weather Observations and Services". These objectives are:

1) The use of solar observations (extreme ultraviolet images, X-ray observations, radio emissions) and models (magneto-hydrodynamic models of flux tubes) for predicting energetic particle events;

2) The use of solar observations and models for predicting coronal mass ejections (CMEs);

3) To develop a quantitative model of the interaction of solar energetic particle events with the Earth's magnetosphere;

4) To develop a quantitative model of the development of trapped radiation in the Earth's magnetosphere during geomagnetic storms;

5) To develop a quantitative model of the variation of galactic cosmic radiation in response to solar activity;

6) To study how electronic technology in satellites, launchers and aircraft is affected by the Earth's radiation environment;

7) To study how humans (astronauts, aircrew, air passengers) are affected by solar and cosmic radiation in different activities;

8) To develop a quantitative model of the propagation of CME through the interplanetary medium to predict their arrival at Earth; and

9) To develop a quantitative model to predict geomagnetic storms and ionospheric current systems.

In 2007, the COST 724 action established the European space weather network 'European space weather web portal on-line' (see under http://gauss.oma.be/COST724/ESWWS/). The COST 724 action intensively cooperates with the COST action 296 "Mitigation of Ionospheric Effects on Radio Systems (MIERS)" [17].

A further European space weather related activity is DIAS (The Digital upper Atmosphere Server). DIAS is co-funded by the e-content programme of the EU and uses real-time information provided by seven operating digital ionospheric stations (ionosondes). DIAS is also based on historical data collections and has developed a pan-European digital data collection on the state of the upper atmosphere over Europe. The first release of the DIAS server prototype is now active on http://www.iono.noa.gr/dias/.

Regional Warning Centres (RWCs) of "The International Space Environment Service" (ISES) deliver tailored space weather forecasts within their own regions [18]. At present there exist seven RWCs, associated RWCs and collaborative expert centres in Europe. These centres are located in the following countries – in Belgium (Brussels), Czech Republic (Prague), France (Toulouse) , Netherlands (Noordwijk, ESA), Russia (Moscow), Poland (Warsaw) and Sweden (Lund). RWC Warsaw and partially RWC Prague inform users about ionospheric forecasts. The Space Weather Application Center – Ionosphere (SWACI)

is a research project at DLR in Germany, which provides nowcasts and forecasts of the ionospheric state to improve the accuracy and reliability of impacted communication and navigation systems for the European user community on a routine basis [19].


**Figure 5-14: SWACI has 10 Partners from Industry (Allsat Hannover, 1A Greifswald),
Governmental (like the Federal Agency for Cartography and Geodesy and others)
and Scientific Organisations and Institutes (like ESA (SWENET), Space
Environmental Center (SEC) Boulder, GFZ (Potsdam and others)).**

On a different level several meetings, workshops and conferences took place, for instance in Armenia, Belgium, Denmark, Finland, France, Germany, Italy, Sweden, UK and other countries. For example in Germany two workshops were organised in 2000 [20] and 2005 [21]. The first workshop delivered a space weather recommendation to industry and organisation in Germany. At the 2005 workshop the space weather community in Germany proposed activities to German authority for a space weather satellite with international cooperation. Discussions on the equipment are directed towards a space weather service related to users from satellite navigation, telecommunication and aviation industry. Especially during the Solar Energetic Particle (SEP) conference in Armenia (2005) American, Armenian, European and Russian space weather experts discussed the important role of cosmic rays for a forecast and nowcast space weather service [22].

5.3.2 Technology, Infrastructure and Space Weather

Parts of many technological systems and infrastructure are strongly influenced by space weather. This is the case for civilian as well as for military systems. Space weather effects are known from the following technological and infrastructure branches:

5.3.2.1 Space and Spacecraft

All spacecraft, both military and civilian, for communications, navigation, remote sensing or research can be affected by the space environment which depends on space weather conditions. The environment a

spacecraft encounters depends on its orbit. For instance, high orbiting GPS and Galileo satellites can be disturbed by rapidly changing conditions in the nearby interplanetary space – by bursts of energetic particles from the Sun (so-called solar particle events SPE with up to several billions of electron-Volts) or Earth orbiting spacecraft may travel through the energetic particles of the Earth's radiation belts or the ionised gases of the upper atmosphere. SPE can have serious consequences for the spacecraft systems for instance:

1) Damage to sensitive spacecraft electronics;

2) Severe degradation of solar panels, thus reducing the area available to collect energy; and

3) Degradation of IR, optical and UV detectors, making them less efficient over time.

Onboard systems can be strongly affected during the solar cycle, especially long-lived satellites. Short-term effects due to space weather storms can trigger communication problems as well as memory malfunctions and data loss. In extreme cases as observed in the 1990s entire satellites were lost.

Strong solar flares trigger geomagnetic storms, which cause heating of the upper atmosphere. Heating of the upper atmosphere causes spacecraft in low Earth orbits to suffer increased drag. In those cases satellite operators have to make corrections for increased atmospheric drag in a timely fashion to avoid loss of the satellite. Contact was lost with the Japanese satellite ADEOS-2 during the intense space weather storm on 24 October 2003 (indirectly by strong enhanced drag). The US Skylab station crashed spectacularly in the southern hemisphere during July 1979 after increasing solar activity caused the atmospheric drag to become stronger than originally expected.

Man-made space debris as an artificial space environment effect is already monitored from US and European organizations. Satellites on low Earth orbits came into contact with rocket stage breaks up in size between microns and meters. The smallest particles frequently impact spacecraft causing damage such as solar cell degradation. Just during the retrieval of LDEF satellite in January 1990 a baseball-size piece of debris flew between LDEF and the space shuttle.

High energy particle radiation is a threat to electronics and other systems. This particle radiation plus high energy electromagnetic X-rays and gamma rays potentially have the ability to break DNA and cause lasting damage to astronauts. Higher levels of shielding in the ISS, or cancelling of extra-vehicular activities helps to minimize radiation exposure to astronauts, but during interplanetary manned missions prolonged exposure leads to large radiation doses and risk of cancer.
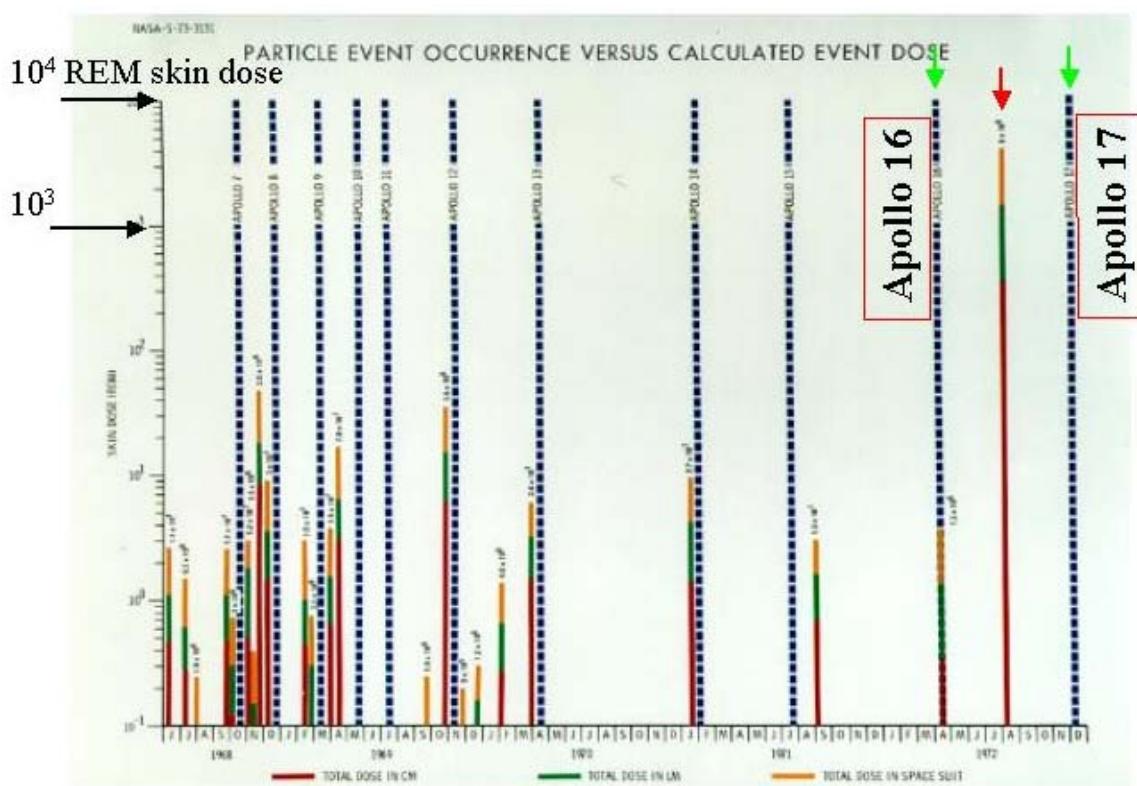
Figure 5-15: SPEs between 1968 and 1973.

It was fortuitous that the large SPE events did not occur whilst the astronauts were travelling to the moon. Had the astronauts been travelling to the moon during an SPE, they may have experienced severe health problems.

Conclusions: Damage to unmanned spacecraft by space weather is not completely avoidable, but plays an important role for high technology networks of infrastructure. Related risks must be recognised, analysed and quantified. Avoidance measures employed by the manufacturers and operators must be closely scrutinised.

5.3.2.2   Satellite Navigation and Telecommunication

Space-based Global Navigation Satellite Systems (GNSS) such as GPS, GLONASS and GALILEO consist of fleets up to 30 satellites. The orbits cross the outer Earth radiation belt at a distance of about 20000 km. In addition to cosmic radiation damages, the most important space weather effect is the degradation of the signal quality, which causes positioning errors making the GNSS less reliable. GPS (U.S. system operational since 1995) and GLONASS (Russian system operational since 1996) were primary military systems, but later partially opened for civil purposes. Space weather storms already impacted the navigation capabilities of both GNSS by:

1)   Interference of GNSS signal strength due to strongly enhanced solar radio emission (MHz and GHz frequency range); and

2)   Delayed signal propagation and degradation of signal performance due to the dynamic behaviour of the ionosphere during space weather storms.

The Galileo satellites will be equipped with radiation monitors.

Solar radio waves and signals of GNSS satellites have frequencies above 30 MHz and are both modified in phase, amplitude and polarization during their ionospheric propagation by fluctuations of the total electron content and other charged particles in its phase, amplitude and polarisation. The resulting range, phase and signal strength errors cause difficulties in the reception of satellite signals for civil and security applications.

Applications such as the detection of minefields, search for mineral resources and oil and gas exploration by means of magnetically guided drilling are partly based on aeromagnetic surveys to map magnetostatic anomalies and rely on an undisturbed geomagnetic field. In addition they often rely on exact positioning by satellite navigation systems. The delivery of high quality data from weather satellites undisturbed by space weather played a major role during Desert Storm in 1991. Measurements and data collecting made during disturbed space weather stormy times result in low-quality results. They must either be repeated or result in missing the target.

Aspects of civil and military telecommunication related to space weather are:

1) The sensitivity and range of radar signals because of changing reflection characteristics in the ionosphere;

2) Ionospheric effects can lead to interference, i.e. signal degradation and delayed echoes;

3) Interference of mobile phone radio waves with solar radio waves in case of solar radio bursts leads to a minimum of failure rate of about 8 % of the calls; and

4) Telecommunication mobile networks working with high capacity are overloaded by extraordinary additional requests of users under space weather storm conditions.

Conclusions: Functioning satellite navigation and telecommunication systems not only represent several billion Euro in investment, but also the potential for damage to the operator's and western society reputations in the event of a failure. Many parts of western society depend on precise satellite navigation and reliable telecommunications. Therefore those service providers must introduce space weather forecast and nowcast information to reduce the navigation and communication problems.

5.3.2.3   Aircraft and Air Traffic

Secondary cosmic rays are produced in the Earth's atmosphere by galactic and solar cosmic rays and directly expose aircrew, passengers and electronics. Flight route, time of flight and aircraft type all influence the radiation exposure. During a flight between the U.S.A. and the Middle East under quiet space weather conditions pilots and military staffs obtain the equivalent of two chest x-rays as an additional effective dose. US and European studies carried out in commercial aircraft found effective dose rates experienced by aircrew between 1 and 15 µSv/h. For military jets and troop carriers similar magnitudes can be adopted or assumed as criteria. A flight on the same route during a SEP can increase the effective dose by an order of magnitude. Space weather effects by galactic cosmic radiation on civil aircrew have been the subject of European Union legislation since 2000. This EU regulation recommend that the annual radiation dose received by aircrew should fall within the 1 – 6 mSv range and that pregnant female aircrew should not be exposed to more than 1 mSv per year. On the other side during large SPEs, cosmic ray dose rates in the nearby Earth environment can increase by a factor of between 100 to 1000! SPE originated radiation dose on cruising altitude of aircraft are: 2.27 mSv (23 February 1956), 1.28 mSv (29 September 1989) and 0.078 mSv (15 April 2001) are much higher dose than about 0.05 mSv from galactic cosmic rays.
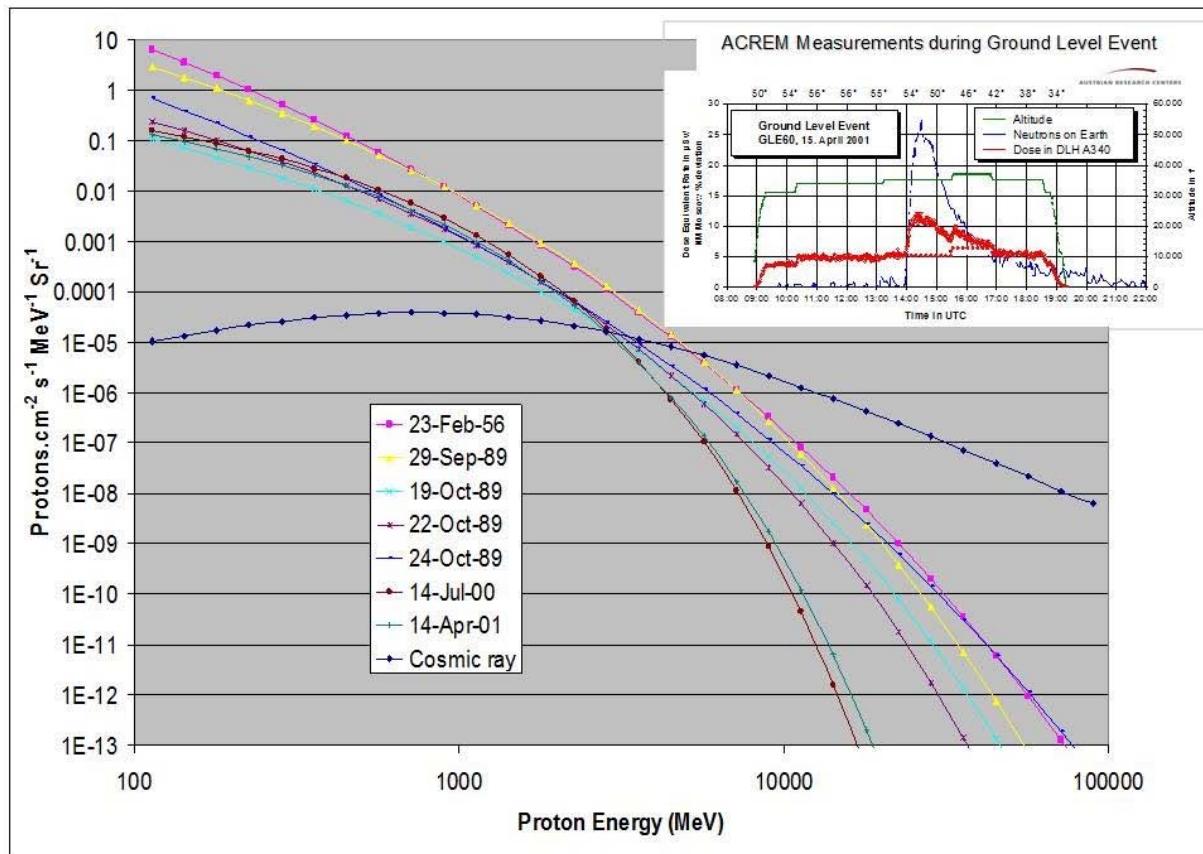
**Figure 5-16: Cosmic Ray Spectra During Major SPEs (C. Dyer, QinetiQ) and
Detection of an SPE on a Cruising Plan (P. Beck, ARCS).**

SPEs let increase the dose rate during a North Atlantic flight to an amount of the same order of magnitude as that allowed for an entire year in Europe. In the U.S.A. the DOT / FAA has also developed cosmic radiation protection standards for air carrier crew members. In principle European and US civil radiation protection recommendations and standards are applicable to military flights. Shielding for sensitive parts and components is not a solution. Any additional shielding increases the mass of the aircraft. Re-routing the flights to lower altitude or to lower latitudes during space weather storms would decrease the cosmic ray exposure, but may be difficult due to logistic reasons and would also increase the costs of flights.

In addition to the dramatic increases seen during SPEs, average cosmic ray doses vary with the solar cycle, because more galactic cosmic rays are able to reach Earth during periods of low solar activity. At the solar minimum between 2007 and 2008, the hourly dose rate was expected to increase by a factor of between 10 % to 50 % compared to solar maximum conditions.

Another critical aspect is secondary cosmic rays potentially interfering with many aspects of the air traffic management system, including communication, navigation and avionics systems in military and civil aircraft. All modern aircraft have onboard electronics that contain microchips operating at very low voltages. Cosmic ray particles can deposit charge in these microchips leading to unexpected changes in the state of memory cells. This can result in unexpected commands or false readings. This type of error is called a single event upset. Software failures and ensuing hardware failures in aircraft systems are documented in scientific journals. Research into the effects of cosmic rays on microchips as well as meetings with pilots have shown that the effects of cosmic rays can be seen as frequently as every two hours at aircraft altitudes. Although the detection of a single cosmic ray is unlikely to result in failure of

any system, this highlights the need for redundancy in those systems (twofold) that are critical for safety, especially for combat zone transport aircraft and jet fighters in operation. Upsets occurred in airplanes of different scale and types – in Boeing, Airbus, private Learjets or others.
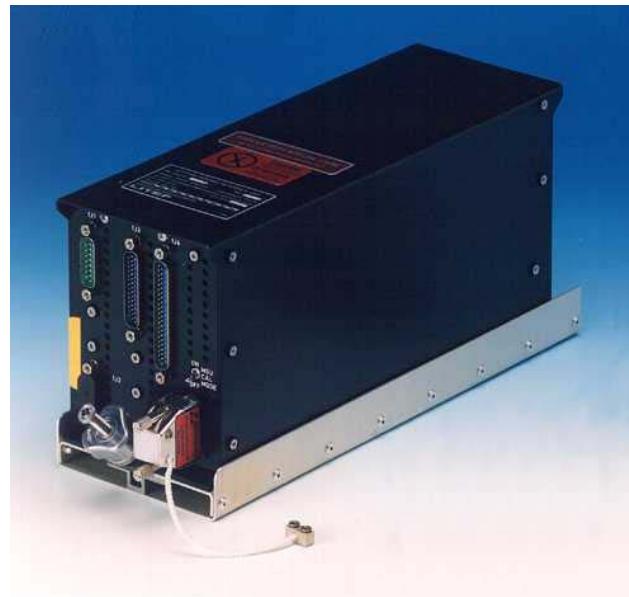
Figure 5-17: The LCR 93 System is in Use for Altitude Information in CESSNA EXEL and LearJet Planes. After cosmic ray induced software failures, new software was developed to compensate for space weather effects (Image: LITEF).
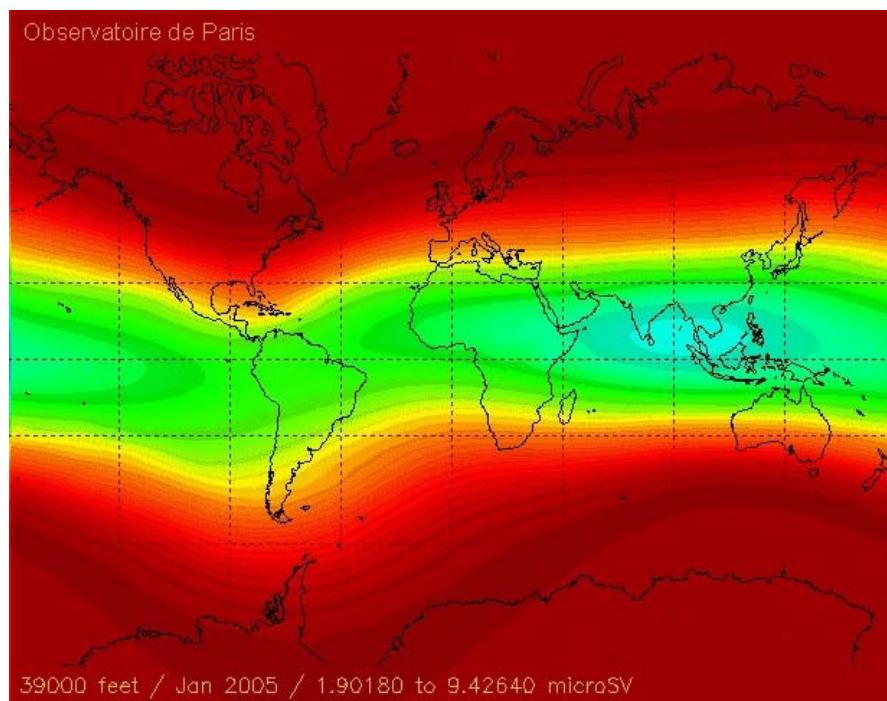

Figure 5-18: Worldwide Map of Estimated Dose Rate (Paris Observatory) during a Solar Particle Storm: Flights near the equator are well protected (white and light green); the maximum dose is obtained near the poles (red).

A third critical aspect is the aircraft communication during space weather events. Solar x-ray flares or energetic particles precipitating from the solar wind modify the lower part of the ionosphere thus changing the propagation conditions for radio waves. This strong sensitivity makes HF communication unreliable in particular at high latitudes. International operations must be aware about risks during high latitude (>50°N) flights.

One advantage aviation has over space flight is that faulty components on a plane can be replaced on the ground and cockpit crews are trained in managing emergency situations.

Conclusions: For security applications the most important potential failures due to space weather are effects on aircraft electronics and communication quality. Regulations for protection of civil aircrew against cosmic radiation exposure became standards in Europe and the U.S.A., however for military flight personnel it has to be studied. Especially in the case when military flight personnel changes to crewmembers of civil airlines the total personal dose has to be known – including radiation dose obtained during flight hours.

5.3.2.4   Power, Oil and Gas Pipeline Supply

Time variations of magnetospheric and ionospheric currents (caused by space weather storms) are seen as geomagnetic variations or disturbances at the Earth's surface. Geomagnetic variations are accompanied by geoelectric fields on the ground. The geoelectric field drives currents in conductors at the Earth's surface, for instance in electric power transmission systems. This current is known as a Geomagnetically Induced Current (GIC). Usually GIC vary much from site to site in a power system. In general, transformers located at corners and ends of a power system suffer from large GIC values and tend to heat up. Long transmission lines also carry larger GIC. How GICs affect the transformer and the whole power system depends on the transformer type, on the structural details of the network and its equipment, on the load and capacity of the power system. GIC are not only a high-latitude problem, although most of the studies performed for power systems in North America and Scandinavia. GIC studies have also been conducted in mid-latitude and equatorial countries, such as South Africa, Brazil and Vietnam.

The largest GIC event so far was experienced on 13 March 1989: A collapse of the Canadian Hydro-Québec power systems left six million people without power for nine hours in winter. The collapse time of Hydro-Québec was only 90 seconds! The absolute value of the changing magnetic field was 400 nT/min. In Minnesota the value was 865 nT/min. The total costs of the Hydro-Québec event were estimated at about 13.2 million Canadian dollars. During the same geomagnetic storm, a high-voltage transformer was completely destroyed in New Jersey. According to Metatech company, the U.S.A. were close to a total power blackout due to cascade effects originating from this super storm. The super storm influenced six 130-kV power lines in Sweden. GICs measured in Finnish power lines had about 130 amps. During the Halloween storm in October and November 2003 power supply failures occurred in South Africa and in Malmö / Sweden: A meltdown of a transformer was attributed due to the intense solar storm induced GIC in South Africa and a hour-long power outage occurred for 20000 inhabitants in Malmö. These and other, nuclear power plants in the U.S.A. and Sweden were discussed in the Department of Homeland Security Infrastructure Report [25] and in the DoC report on Halloween storms in 2003 [26]. The great power black outs in summer 2003 in New York, London and Rome were not caused by space weather storms, but showed how infrastructure systems and millions of people depend on a stable energy production!
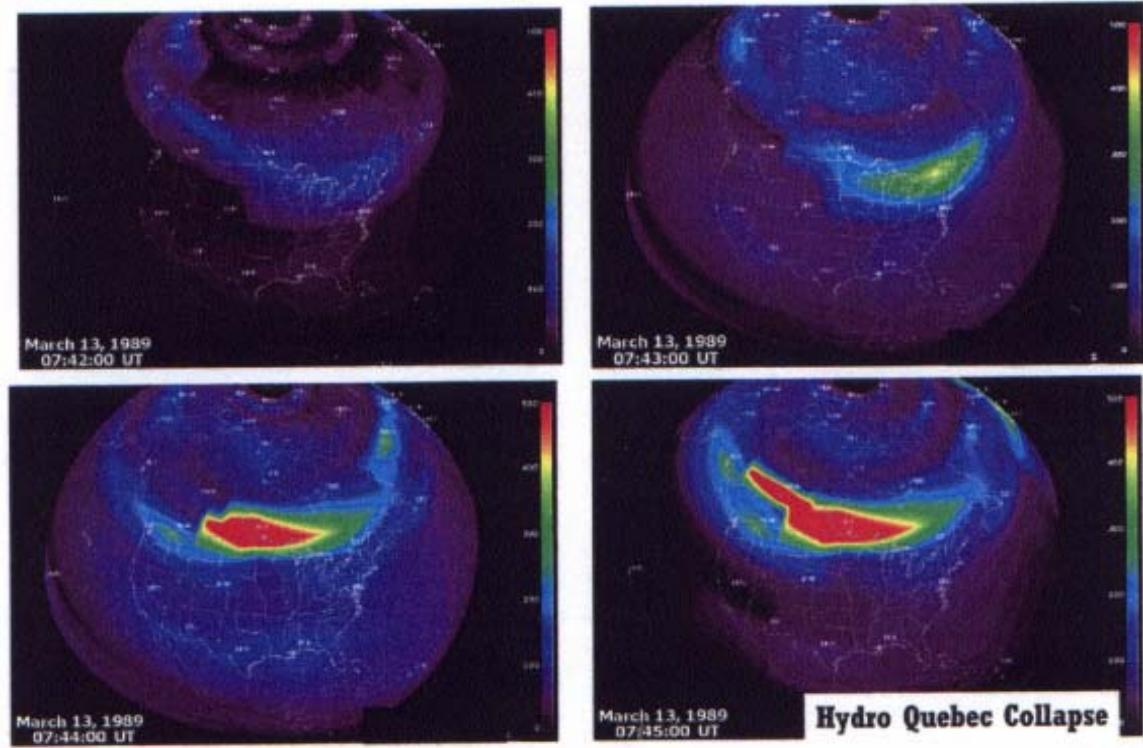
Figure 5-19: March 1989 – The Super Space Weather Storm Arrived at the
US East Coast and Moved to the West Coast Within Four Minutes.

Figure 5-20: High-Voltage Power Transformer in New Jersey, USA, Before and After the
Geomagnetic Super Storm (first and second image) on the 13th March 1989 and
the Transformer Windings were Permanently Damaged in New Jersey,
USA, During the March 1989 Geomagnetic Storm. Third image:
destroyed transformer in South Africa in October 2003.
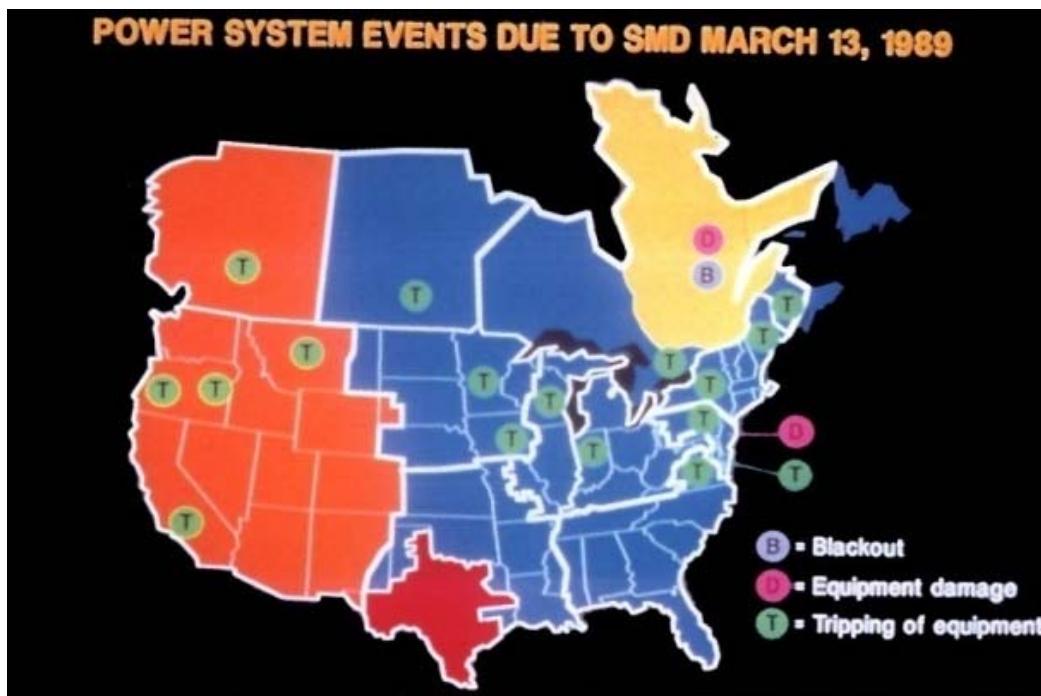
**Figure 5-21: Map Showing Many Problems Caused by the March 1989
Geomagnetic Storm in North American Power Systems.**

Besides power grids, GIC also flow in oil and gas pipelines. To avoid corrosion, pipelines are covered by an insulating coating and equipped with a cathodic protection system to prevent electric currents from the pipeline. Amplitudes of pipe-to-soil voltages related to GIC can easily exceed the cathodic protection potential thus removing the protection. It is unknown how strongly GIC effects increase the corrosion rate of a pipeline, because space weather risk has not been investigated as much in pipelines as in power systems. However GIC studies and GIC induced measurements on pipelines have been carried out in Alaska, Canada, Finland, Germany and Sweden.
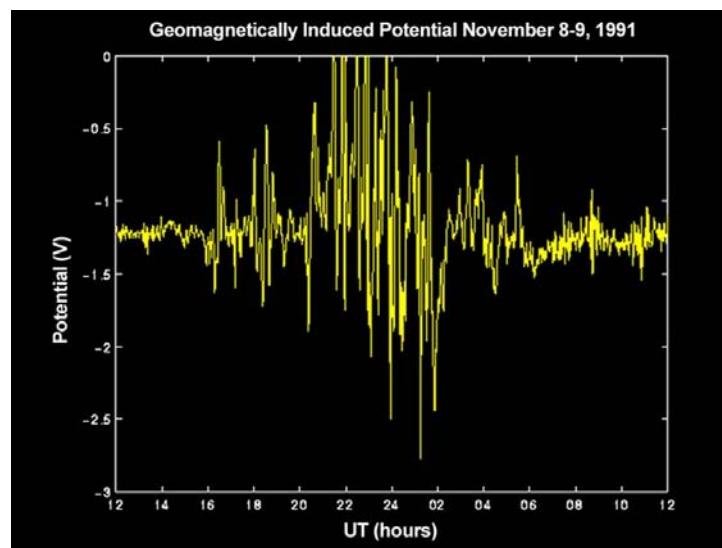
**Figure 5-22: GIC-Related Pipe-to-Soil Voltage Variations Measured in the Swedish Pipeline on
8-9 November 1991. Note that the zero level is about -1.2 V indicating
the value of the cathodic protection voltage.**
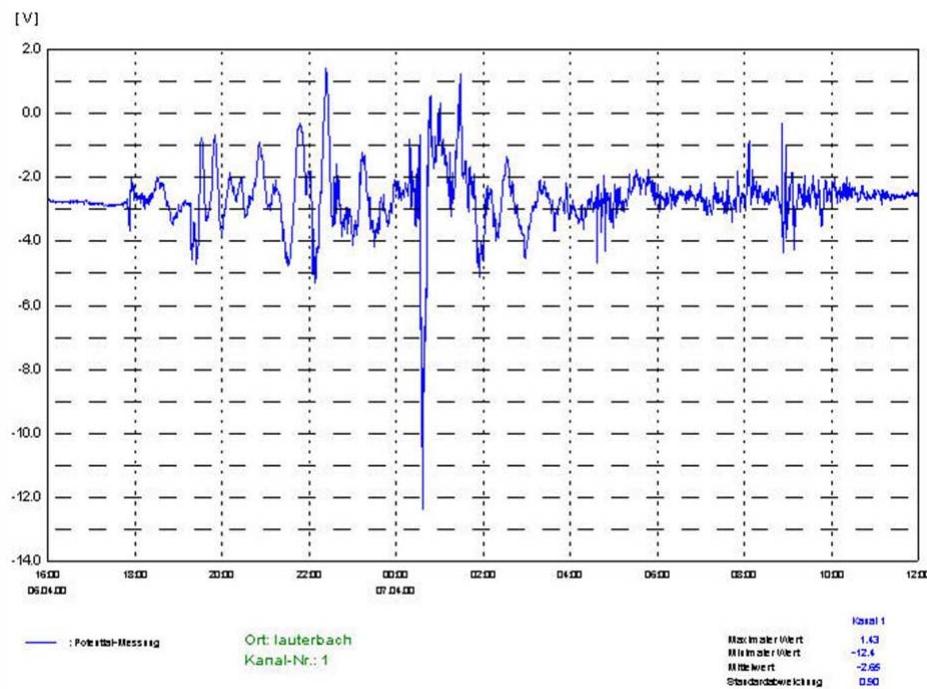
**Figure 5-23: The Ruhrgas Pipeline in Germany is a Good 'Detector' of GIC – 6-7 April 2000.**

Conclusions: Space weather storms have induced power cuts in the U.S.A., Europe and South Africa. Such events demand a regional and global risk analysis, a risk strategy and risk control in case of big solar storms.

5.3.2.5   Telecommunication Cables, Railway Equipment and Electronics

GIC also flow in other technological networks, such as telecommunication cables and railway power lines and traffic equipment, causing problems to the normal operation of the systems affected. The use of electrical telecommunication cables is decreasing in the developed countries. However these cables are still affected by GICs during space weather storms. Space weather has disrupted electric lines, railway signalling systems and engine drives as reported below. The first know example was a disruption of railway traffic in New York City on 15 May 1921.

In July 1982 railway signals switched from green to red light and vice versa due to strong geomagnetic activity near Stockholm in Sweden. In the 1990s the German high speed train ICE was equipped with high power semiconductors (GTO) in the engine drive. About 50% of all trains were affected by space weather. Cosmic ray particles produced electron bursts in layers of the semiconductors, completely destroying the layers. Therefore ICEs were driven with much smaller speed than the nominal 250 km/s during cosmic ray bursts. Later ICE equipment – so-called IGBTs – seems to be less affected. IGBT are also used in subways and trams. Significant links between geomagnetic disturbances and anomalies in the operation of railway automatics and disturbances occurred in 2004 / 2005 in the East Siberian Railway.

Electronic components form a vital part of today's western societies. Everything from cars, trucks to a high-tech aircraft and spacecraft contain electronics. As technology advances, the components become smaller and smaller. This miniaturisation allows computers to operate at much lower voltages than before. As a result, circuits are more susceptible to additional currents created by the impact of cosmic ray particles. Energetic cosmic ray air shower particles, created through the interaction of cosmic rays with the Earth's atmosphere, are capable of disrupting electronic systems in space, air and on the ground. In 1978,

IBM experts first studied the effects of secondary cosmic rays on memory chips in PCs [27]. The software errors caused by the interaction of the particles with memory devices, however, should not affect the normal operation of a PC because the system is developed such that it can detect the error and compensate for it. However cosmic ray induced errors in critical car and truck electronics are discussed in the industry. For instance significant implications for mission-critical automotive applications that utilize a so-called SRAM-based FPGA seems to be established.

Conclusions: Several important infrastructure and transport systems failures due to space weather have been reported. Modern societies depend on electronics in transport systems, i.e., railways, subways, trams, cars, aircraft, spacecraft and ships. The security of those systems is a foundation of a reliable infrastructure in modern societies and must be safeguarded against space weather failures. Several other effects to similar technical systems may have already occurred, but the producers, owners or users are often not aware of the space weather background of failures. Further studies, education and development of special action plans to diminish the effects of failures are very good measures.

5.3.3   Insurance

Space industry insurance experts estimate that over USD 500 million of the total USD 4,000 million losses can be linked to space weather in the second half of the 1990s. Over 180 satellites were insured in the same period for USD 24,000 million. Another 200 were not insured. These were larger research satellites and military satellites. In the first decade of the 21st century 1,400 satellites with a value of USD 220,000 million are planned to be launched. Of those, 900 will be used for commercial purposes. Considering that safety measures on these satellites are kept to a minimum in order to keep costs down, potential losses will increase significantly by the next solar maximum after 2011.

Investment by civil and military spacecraft operators in space weather services could provide one way in which the amount of shielding required could be reduced without increasing the risk to the spacecraft.

Many more risk reduction measures can be implemented in aviation. Increasingly precise early warning or forecasting systems now make it possible to change flight paths, thereby reducing the exposure of passengers, crew and the plane's technological systems.

Satellites and airplanes usually have comprehensive insurance. However, those responsible for space flight and air travel must be aware of space weather phenomena, taking organisational and technical measures to protect flight personnel and electronics from excessive exposure to high levels of cosmic radiation.

A remark to ground-based effects: It is far easier to assume that a transformer was destroyed by a technical defect than to imagine that space weather might have caused the loss. From the perspective of the insurance industry as a whole, lengthening the chain of cause and effect is not difficult. For example, a transformer fire such as in New Jersey in March 1989, can cause property loss and also lead to business interruptions. Insurance policies do not generally distinguish between fires caused by lightning, technical failure or a GIC. Experts estimated the economic loss in the New Jersey case to be USD 3,000 to 6,000 million. Only a small fraction of this was insured. Technically, this consisted of losses incurred within the company following the transformer fire, and losses incurred by a third party, for example, as a result of the power cut.

The following general factors are important for space weather insurance aspects:

- Increasing interdependency, particularly in the area of electrical and electronic systems (e.g., communication);

- The marked increase in systems liable to fail (although a certain degree of risk is accepted in the interests of cost saving);

- The complex relationship between production and communication processes, whose interdependencies multiply in the event of a loss; and

- Air travel: the rapidly growing importance of polar flight paths, e.g. from the U.S. to Asia (the dose encountered during a New York – Hong Kong flight is some 0.2 milliSieverts).

Variability in space weather means that the probability and extent of losses which may be incurred by the various companies involved in space and air travel, electric power generation, telecommunications or oil and gas transport, are not necessarily comparable. For this reason, individual actuarial assessment must be performed for each insured company, so that the most appropriate insurance premium can be calculated.

It is the insurance industry's responsibility to provide information on and raise awareness of space weather. The challenge is to develop, adapt and bring to the market products to insure the diverse range of potential losses related to space weather (see in [28]).

Conclusions: It is the responsibility of the insured organization or persons to implement risk-mitigating measures. Therefore early warning systems capable of detecting extraordinary solar activity and space weather storms - accurately and in good time - will become crucial in the future for insurance industry and infrastructure systems. In order to retain insurance cover, individuals may be required to take cognisance of any forecasts.

5.3.4   Space Weather and Security Aspects

In addition to the directly affected space weather technology and infrastructure described in chapter 2, several studies have been carried out relating to security aspects. Examples for overlapping civil and security aspects were the nuclear power plant and power grid problems in October / November 2003 in the U.S.A. and in Sweden. These were discussed in the DoC and DHS reports ([25], [26]).

HF satellite communication can be interrupted between half an hour and several hours due to large solar flares. One dangerous example took place during the Gulf war in 1991 when requests for fire support were made by HF radio communications, which were disrupted.

Over the Horizon radars (OTH) utilises HF propagation of signals over very large distances up to thousands of kilometres. The waves propagate in a duct between the ionosphere and the ground. OTH radar propagation can be enhanced either naturally or artificially. Natural enhancement is the result of space weather storms by CMEs and the influence of magnetic storms. Artificial enhancement can be achieved by locally heating the ionosphere, thus increasing its conductivity at the reflection point. High energies are needed to achieve this. By heating the local ionosphere using strong radio power, space weather effects may even be imitated to disrupt the radio communication and navigation services of the enemy.

**Figure 5-24: A High Power Antenna Field in Alaska
(with a 'Dust Devil' in Background (image: F. Jansen)).**

Military radio systems are sensitive to space weather events, mainly due to the ionospheric impact on radio signals. Radar systems may suffer from limitations caused by the variability of the ionosphere. This is especially true for the detection of low flying targets or moving ships. Here the detectability of the target is reached by summing successive echoes at each rotation of the radar antenna. Ionospheric effects can lead to interference, i.e. signal degradation and delayed echoes. Since military users need reliable communication and precise navigation, ionospheric space weather effects may cause serious problems in the detection of aircraft and ships. Studies with radars at the east cost of Canada and the north cost of Australia showed that ship detections is strongly depending on the status of the ionosphere.

The French Nostradamus is an over the horizon radar demonstrator, capable of detecting and locating an aircraft well behind the horizon – the limit for conventional radars. Emitting very low frequency waves (6-30MHz) that bounce off the ionosphere, Nostradamus can transmit and receive signals over a range of over 1000 km.
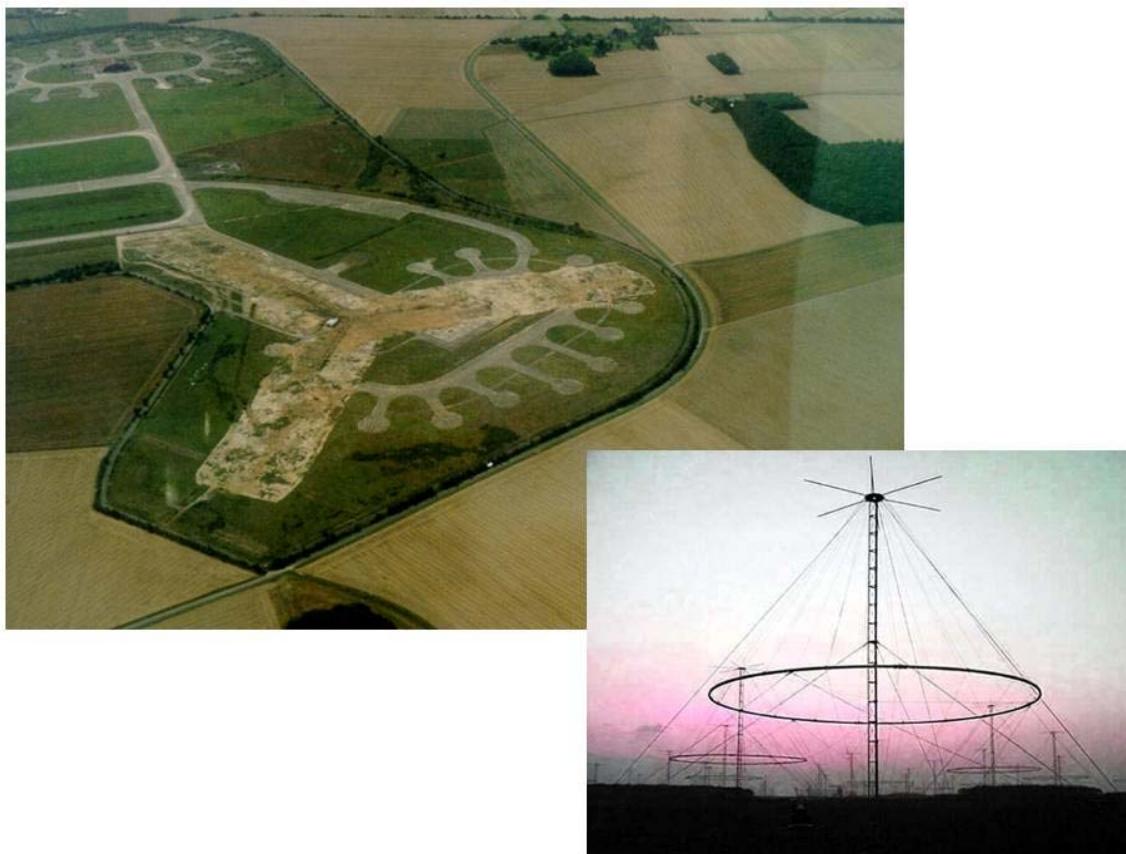
Figure 5-25: Aerial View and Antenna of the Nostradamus OTH Radar Located West of Paris. Each of its 3 arms stretches several hundred metres and comprise 100 antennas.

Different security-related activities of space weather in Europe have initiated or are in progress. Within the ESA space weather feasibility studies British and French Armed Forces were contacted by the Alcatel and RAL team members.

Later within the COST action 724, National Representatives contacted the German Bundeswehr, the Spanish Defence Ministry and the Swedish Defence Research Agency (FOI sponsored by the Swedish Armed Forces SWAF). The German Bundeswehr announced interest in the potential usage of MuSTAnG space weather telescope data, including access to the data from the international muon telescope network to predict 10 – 20 hours in advance the onset time of communication disturbances in the ionosphere. The German Bundeswehr is already using ionospheric data from Leibniz-Institute of Atmospheric Physics in northern Germany. FGAN – Research Establishment for Applied Sciences – in Wachtberg / Germany is carrying out space weather related ionospheric research, too. Activities include the correction of propagation-dependent bearing estimation errors of HF radio waves. The DG of INTA (Spanish acronym for National Institute of Aero-Space Technology www.inta.es) evaluate the national interest in the space weather topic and INTA would handle any future action. Members of the Swedish Defence Research Agency FOI attended US space weather week, discussed ESA space weather programs, EU COST action 724 and wrote a report. In the report it is recommended to create a unit within the Swedish Armed Forces to deliver space weather information to the divisions of the Armed Forces. FOI also recommend courses about space weather and effects for the Armed Forces in Sweden. According to Swedish National Representative it is very important for SWAF to have knowledge about space weather and its effects that are being used and owned by the SWAF during international operation.

A recent study done by SEA Ltd. / UK have shown the benefits of a 50/50 split between nowcast and forecast tools for a civil European Space Weather Service (ESWS). Furthermore, small initial investment will bring large benefits – to a civil as well as a security related space weather service.

A highlight of the European efforts will be a Space Situational Awareness (SSA) program, which is presently in preparation for adoption at the ESA Ministerial Conference in November 2008. It is foreseen that the optional SSA program will cover navigation, communications and reconnaissance activities. Especially it will focus on topics like space weather, near-Earth objects, as well as space debris. Main proposed deliverables are a radar array, and one to two LEO or GEO satellites. The SSA budget may reach about 500 million Euro from 2009 to 2019 and may be a cornerstone for security aspects.

Conclusions: At present and much more in future, civil and military daily life rely heavily on high technology. The idea that technological systems could be threatened by hazards originating beyond the atmosphere is not strange anymore for many organizations and people. Public, political and security awareness on space weather has rapidly increased.

5.3.5   References

[1]   Jansen, F., and R. Pirjola, 2004, Space Weather Research Elucidates Risks to Technological Infrastructures, Eos 85 (25), 242-246.

[2]   Jansen, F. (ed. for EU / SWEETS consortium), 2007, Space Weather DVD, 8 language version.

[3]   Daly, E., Crosby, N, and Hilgers, A (eds.), 1999, Workshop on Space Weather, ESA/ESTEC, WPP-155, ISSN 1022-6656, 1-528.

[4]   Jansen, F., Hippler R., 2003, Internal Reports submitted to ESA.

[5]   Daglis, I.A. (ed.), (2001), Space Storms and Space Weather Hazards, Proc. NATO ASI, NATO Science Series II: Mathematics, Physics & Chemistry, Kluwer Academic Publishers, ISBN 1402000316, 1-492.

[6]   Solar Cycle and Space Weather (SOLSPA), Proc. Euroconference, ESA-SP-477, 2002.

[7]   Workshop on The Utilization of a Future European Space Weather Service 12 December 2000, ESTEC Noordwijk, The Netherlands.

[8]   Proceedings of the Space Weather Workshop: Looking Towards a Future European. Space Weather Programme, 17 – 19 December 2001, ESTEC, Noordwijk, The Netherlands, ESA WPP-194.

[9]   ESA Space Weather Workshop: Space Weather Applications Pilot Projects 16-18 December 2002, ESTEC Noordwijk, The Netherlands.

[10] ESA Space Weather Workshop: Developing a European Space Weather Service Network, 3-5 November 2003, ESTEC Noordwijk, The Netherlands.

[11] Jansen, F. and Hippler, R., (eds. for EU / SWE consortium), (2003) Space Weather CD-Rom, 3 language version.

[12] Proc. 1st European Space Weather Week under http://www.esa.spaceweather.net/spweather/workshops/esww/proceedings.html, ESTEC Noordwijk 29 November – 3 December 2004, The Netherlands.

[13] Proc. 2nd European Space Weather Week under http://www.esa-spaceweather.net/spweather/workshops/eswwII/esww2-proceedings.html, ESTEC Noordwijk 14-18 November 2005, The Netherlands.

[14] EU/ESA White Paper on http://ec.europa.eu/comm/space/whitepaper/whitepaper/whitepaper_en.html.

[15] COST 724 action on http://cost724.obs.ujf-grenoble.fr/representatives.html.

[16] COST 724 space weather school Trieste on http://cost724.obs.ujf-grenoble.fr/index-welcome.html.

[17] COST 296 action on http://www.cost296.rl.ac.uk/.

[18] ISES under http://www.ises-spaceweather.org/about_ises/index.html.

[19] SWACI under http://www.kn.nz.dlr.de/swaci/.

[20] 1st National Space Weather Workshop at DLR / Neustrelitz under http://www.kn.nz.dlr.de/space_weather/workshop/haupt.html, 2000, Neustrelitz.

[21] 2nd National Space Weather Workshop at DLR/Neustrelitz under http://www.kn.nz.dlr.de/Topics/Space Weather, 2005, Neustrelitz.

[22] Jansen, F., Solar Extreme Events (SEE-2005) International Symposium at Nor Ambered, Armenia, in press.

[23] Kiernan, Vincent, 'Satellites Play Key Role in Swift Gulf Victory', Space News, 4 March 1991, page 1, 20.

[24] Jansen, F., 2001, Space Weather Reports on www.weltraumwetterwarte-greifswald.de.

[25] Department of Homeland Security (DHS) Infrastructure Report, November 03, 2003.

[26] DoC Space Weather Storm Service Assessment Report, April 2004.

[27] Ziegler, J.F. et al., IBM J. Res- Develop. Vol. 40 No. 1.

[28] Jansen, F., Pirjola, R., Favre, R., 2000, Space Weather: Hazard to the Earth?, Swiss Re -insurance space weather study under http://www.swissre.com/.

Appendix 1 – PROCEEDINGS FROM IST-056 SYMPOSIUM

To provide a forum for interested parties to present and discuss their research, the technical committee of IST-051 organised a symposium addressing the subject of characterising the ionosphere. The symposium was a natural extension to the work of the task group sponsored by NATO. To broaden the scope, the relevant group from URSI was invited to participate and they kindly accepted. The symposium was held at the Geophysical Institute, University of Alaska Fairbanks, from 12th to 16th June 2006.

The following are the papers which were presented. A link to a copy of the proceedings is provided below; alternatively the proceedings are available under ISBNs 92-837-0078-3 / 978-92-837-0078-4.

Paper 1 – Studies of Ionospheric Processes in the Atmosphere and the Laboratory
Slanger, T.G.

Paper 2 – Investigating the Auroral Thermosphere with N2+ Lidar
Collins, R.L.; Su, L.; Lummerzheim, D.; Doe, R.A.

Paper 3 – Polar Ionospheric Imaging at Storm Time
Yin, P.; Mitchell, C.; Bust, G.

Paper 4 – Morphology of Southern Hemisphere Riometer Auroral Absorption
Foppiano, A.J.

Paper 5 – An Investigation into the Relationship between Ionospheric Scintillation and Loss of Lock in GNSS Receivers
Meggs, R.W.; Mitchell, C.N.; Smith, A.M.

Paper 6 – Morphological Characteristics of L-Band Scintillations and their Impact on GPS Signals – A Quantitative Study on the Precursors for the Occurrence of Scintillations
Rama Rao, P.V.S.; Tulasi Ram, S.; Gopi Krishna, S.; Niranjan, K.; Prasad, D.S.V.V.D.

Paper 7 – Observations of the Tongue of Ionization with GPS TEC and SuperDARN
Coster, A.; Colerico, M.; Foster, J.C.; Ruohoniemi, J.M.

Paper 8 – VLF Phase Perturbations Produced by the Variability in Large (V/m) Mesospheric Electric Fields in the 60 – 70 km Altitude Range
Manson, A.H.; Meek, C.E.; Martynenko, S.I.; Rozumenko, V.T.; Tyrnov, O.F.

Paper 9 – Withdrawn Paper 10 – Ionospheric F-region Storms: Unsolved Problems
Prölss, G.W.

Paper 11 – Space Weather Applications of the UAF Eulerian Parallel Polar Ionosphere Model (EPPIM)
Maurits, S.; Kulchitsky, A.; Watkins, B.

Paper 12 – Theoretical and Observational Studies of Meteor Interactions with the Ionosphere
Colestock, P.; Close, S.; Zinn, J.

Paper 13 – European Space Weather Activities
Jansen, F.

Paper 14 – High Latitude Ionospheric Structures
MacDougall, J.

Paper 15 – Large-Scale Plasma Structure in the Polar and Auroral Ionosphere: Experimental Observations and Modelling
Pryse, S.E.; Middleton, H.R.; Dewis, K.L.; Wood, A.G.; Whittick, E.L.; Balthazor, R.L.

Paper 16 – Direction Finding Errors Induced by Plasmawaves of the Ionosphere
Hawlitschka, S.

Paper 17 – Characterisation of Narrowband HF Channels in the Mid and Low Latitude Ionosphere
Harris, T.J.; Scholz, M.L.

Paper 18 – Quasi-Analytic Models for Density Bubbles and Plasma Clouds in the Equatorial Ionosphere
Bernhardt, P.A.

Paper 19 – Modeling of Ionospheric Refraction of UHF Radar Signals at High Latitudes
Watkins, B.; Maurits, S.; Kulchitsky, A.

Paper 20 – Mitigation of Ionospheric Effects on High Frequency Surface Wave Radar
Riddolls, R.J.

Paper 21 – Ground-Based Radar Detection of the Inner Boundary of the Ion Plasma Sheet and its Response to the Changes in the Interplanetary Magnetic Field
Jayachandran, P.T.; MacDougall, J.W.; Moorcroft, D.R.; Donovan, E.F.

Paper 22 – The European Server for Ionospheric Specification and Forecasting: Final Results from DIAS Project
Belehaki, A.; Cander, Lj.; Zolesi, B.; Bremer, J.; Juren, C.; Stanislawska, I.; Dialetis, D.; Hatzopoulos, M.

Paper 23 – A Digital Radio Receiver for Ionospheric Research
James, H.G.

Paper 24 – Not released for publishing

Paper 25 – What Can We Learn About the Ionosphere Using the EISCAT Heating Facility?
Rietveld, M.T.

Paper 26 – GPS Sounding of the Ionosphere Onboard CHAMP
Jakowski, N.; Mayer, C.; Wilken, V.

Paper 27 – Real-Time Imaging of the Ionosphere over the United Kingdom – Preliminary Results
Meggs, R.W.; Mitchell, C.N.; Watson, R.J.; Dear, R.M.

Paper 28 – Characterization of the Ionosphere over the South Atlantic Ocean by Means of Ionospheric Tomography using Dual Frequency GPS Signals Received on Board a Research Ship
Cilliers, P.J.; Mitchell, C.N.; Opperman, B.D.L.

Paper 29 – GPS Users Positioning Errors during Disturbed Near-Earth Space Conditions
Afraimovich, E.L.; Demyanov, V.V.; Tatarinov, P.V.; Astafieva, E.I.; Zhivetiev, I.V.

Paper 30 – Withdrawn

Paper 31 – Review of the Current Status of Four-Dimensional Ionospheric Imaging
Bust, G.S.; Mitchell, C.N.

Paper 32 – The International Reference Ionosphere – Climatological Standard for the Ionosphere
Bilitza, D.

Paper 33 – Withdrawn

Paper 34 – Modeling of Sporadic-E Structures from Wind-Driven Kelvin-Helmholtz Turbulence
Bernhardt, P.A.; Werne, J.; Larsen, M.F.

To obtain a copy of the RTO-MP-IST-056 proceedings, click on the link below:

RTO-MP-IST-056 Proceedings

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-TR-IST-051 AC/323(IST-051)TP/207	ISBN 978-92-837-0057-9	UNCLASSIFIED/ UNLIMITED

5. Originator
Research and Technology Organisation
North Atlantic Treaty Organisation
BP 25, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Characterising the Ionosphere

7. Presented at/Sponsored by
Final Report of Task Group IST-051.

8. Author(s)/Editor(s)	9. Date
G. Wyman	January 2009

10. Author's/Editor's Address	11. Pages
Little Chase, Tidenham Chase, Chepstow NP16 7JN, UK	290

12. Distribution Statement
There are no restrictions on the distribution of this document.
Information about the availability of this and other RTO
unclassified publications is given on the back cover.

13. Keywords/Descriptors

Auroral and polar cap phenomena	Ionospheric storms
Electron Density Assimilative Model (EDAM)	Ionospheric troughs
GPS	Multi-Instrument Data Analysis System (MIDAS)
HF propagation	
HF radars	Radio tomography
Ionosonde	Satellite communications
Ionospheric irregularities	Space weather
Ionospheric plasma	Travelling Ionospheric Disturbance (TID)

14. Abstract

This report is a compilation of papers from academic and other research institutes that describe ionospheric phenomena and parameters that affect electromagnetic propagation. The report addresses both mid-latitude and high-latitude effects. The first chapter describes the morphology of the ionosphere, briefly covers the neutral atmosphere and expands on the influence of the geomagnetic field, ionospheric electric field and currents, and particle precipitation at high latitudes. Chapter 2 looks at the instrumentation for obtaining the relevant data. An analysis of the mapping of total electron content in geographic co-ordinates is presented in Chapter 3 with an aim of providing timely information to the users. Chapter 4 addresses spontaneous emissions observed during ground-based and rocket-borne experiments. The final chapter considers the specific effects of space weather on the ionosphere during severe geomagnetic and ionospheric storms, the impact on technology including communication and navigation systems, and emphasises the need for a global perspective of the system. Extensive references are provided for each subject. For completeness, the proceedings of a Specialists' Meeting held in conjunction with the study are added as an appendix.

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS

RTO NON CLASSIFIEES

Les publications de l'AGARD et de la RTO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre est la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la RTO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (www.rto.nato.int) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

CANADA
DSIGRD2 – Bibliothécaire des ressources du savoir
R et D pour la défense Canada
Ministère de la Défense nationale
305, rue Rideau, 9e étage
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5, 2750 Ballerup

ESPAGNE
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE (Correspondant)
Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALIE
General Secretariat of Defence and
National Armaments Directorate
5th Department – Technological
Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE
Centralny Ośrodek Naukowej
Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

ROUMANIE
Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

ROYAUME-UNI
Dstl Knowledge and Information
Services
Building 247
Porton Down
Salisbury SP4 0JQ

SLOVENIE
Ministry of Defence
Central Registry for EU and
NATO
Vojkova 55
1000 Ljubljana

TURQUIE
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanliklar
Ankara

AGENCES DE VENTE

**NASA Center for AeroSpace
Information (CASI)**
7115 Standard Drive
Hanover, MD 21076-1320
ETATS-UNIS

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Les demandes de documents RTO ou AGARD doivent comporter la dénomination « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants :

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources
uniformes (URL) suivant: http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme
NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
ETATS-UNIS

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
ETATS-UNIS
(accessible également en mode interactif dans la base de
données bibliographiques en ligne du NTIS, et sur CD-ROM)

BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED
RTO PUBLICATIONS

AGARD & RTO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO reports, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your Organisation) in their distribution.

RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of RTO reports as they are published, please visit our website (www.rto.nato.int) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
   for Defence, National RTO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30
1000 Brussels

CANADA
DRDKIM2 – Knowledge Resources Librarian
Defence R&D Canada
Department of National Defence
305 Rideau Street, 9th Floor
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
LOM PRAHA s. p.
o. z. VTÚLaPVO
Mladoboleslavská 944
PO Box 18
197 21 Praha 9

DENMARK
Danish Acquisition and Logistics Organization (DALO)
Lautrupbjerg 1-5
2750 Ballerup

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY
Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)
Defence Industry & Research General Directorate
Research Directorate, Fakinos Base Camp
S.T.G. 1020
Holargos, Athens

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
P O Box 26
H-1525 Budapest

ITALY
General Secretariat of Defence and
   National Armaments Directorate
5th Department – Technological
   Research
Via XX Settembre 123
00187 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
Royal Netherlands Military
   Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY
Norwegian Defence Research
   Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Centralny Ośrodek Naukowej
   Informacji Wojskowej
Al. Jerozolimskie 97
00-909 Warszawa

PORTUGAL
Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA
Romanian National Distribution
   Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353, Bucharest

SLOVENIA
Ministry of Defence
Central Registry for EU and
   NATO
Vojkova 55
1000 Ljubljana

SPAIN
SDG TECEN / DGAM
C/ Arturo Soria 289
Madrid 28033

TURKEY
Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
   Başkanlığı
06650 Bakanliklar – Ankara

UNITED KINGDOM
Dstl Knowledge and Information
   Services
Building 247
Porton Down
Salisbury SP4 0JQ

UNITED STATES
NASA Center for AeroSpace
   Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace
   Information (CASI)
7115 Standard Drive
Hanover, MD 21076-1320
UNITED STATES

The British Library Document
   Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and
   Technical Information (CISTI)
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa K1A 0S2, CANADA

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource
locator: http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific
   and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
UNITED STATES

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 2216
UNITED STATES
(also available online in the NTIS Bibliographic Database
or on CD-ROM)